DESIGN AND EVALUATION OF A COMPUTER BASED SYSTEM TO MONITOR AND GENERALISE, BY AREAS, DATA FROM ERTS PRECISION IMAGERY TAPES

Professor Keith M. Clayton
University of East Anglia
School of Environmental Sciences
Norwich NOR 88C
England

March 1974

Type II Report for Period 23 July 1973 - 22 January 1974

Department of Trade and Industry
Abell House
John Islip Street
London SW1P 4LN
England

1. SR No. NMC 035	2. Type of Report II	3. Recipient's Catalog No.
4. Title Design and evaluation of a computer based system to monitor and generalise, by areas, data from ERTS precision imagery tapes.		5. Report Date 6. Period covered July 1973 - January 1974
7. Principal Investigator Professor Keith M Clayton		8. No. of Pages
9. Name and Address of Principal Investigator's Organization School of Environmental Sciences University of East Anglia Norwich NOR 88C, England		10. Principal Investiga.Rept.No. 11. GSFC Technical Monitor George J. Ensor
12. Sponsoring Agency Name and Address Department of Trade and Industry Abell House, John Islip Street London SW1P 4LN		13. Key Words (Selected by Principal Investigator
14. Supplementary Notes		

15. Abstract

This is a detailed progress report covering all aspects of the investigation during the perood December 1972 until February 1974. Significant results are presented concerning the accuracy with which Unit Test Areas can be located in CCT coordinates, and the extent to which the small scale structures in the gray-scale histograms are related to real features on the surface. In addition, progress in the development of analytical techniques is presented.

CONTENTS

1. PREFACE

1.1 Objectives

The aim of this project is to devise a system for generalising remotely sensed data for areas of the earth's surface at a scale appropriate to a world data collection system. Successful generalisation will allow the identification of changes in surface cover over successive orbits, and may be used for small scale mapping of these changes.

This system, if satisfactory, could be used as a first stage data filter, selecting for further analysis those images which show significant changes from earlier imagery. It is desirable that the system uses data which has undergone the minimum amount of pre-processing, and hence bulk (system corrected) ERTS imagery has been used throughout.

Since it is likely that any operational earth observation satellite will, like ERTS-1, return the imagery as a digitized picture, there are many advantages in devising a system capable of processing this raw digital data. Among these advantages are the reduction in slow and costly photographic processing, the potential for developing a fully automatic system, and the ability to make use of the full dynamic range of the sensors without the degradation in radiometric fidelity introduced by photographic processing. Also future systems will possibly have sensors which do not yield information suitable for photographic presentation.

1.2 Generalisation of the Imagery

Each ERTS frame contains 3×10^7 pixels (picture elements) for the four MSS bands, whereas a regional map at a scale of, say, $1 : 10^6$ could show $10^2 - 10^3$ independent data points in an equivalent area. It is thus obvious that some generalisation of the ERTS imagery is required for regional scale investigations.

This generalisation can be introduced in two ways. The conventional approach is to classify the imagery into surface types, and to use these surface types as a basis for generalisation. One disadvantage of this approach is the expense in computing power required to classify each pixel quite apart from the difficulty of recognizing 'standard types' for such classification. An alternative system is to generalise the raw data prior to classification and mapping on a regional scale. This second approach is adopted for this project, since it offers the capability of efficiently monitoring the imagery on a regional scale.

The types of feature which should be significant on this scale are, for example, large grass burns in savanna regions, or snow cover in the catchment area of a reservoir. It is hoped that the system described here could identify this type of change against a background noise of smaller scale changes.

1.3 Generalisation Units

One of the aims of the project is to determine the area over which the data can be effectively generalised and yet still give a useful representation of the image. It is probable that this area will be a function of the terrain type and the features of interest.

Initially, square Unit Test Areas (UTAs) covering 2.5×10^5 hectares have been used. These UTAs are defined by the geographical coordinates of their centre points, and have edges parallel to lines of latitude and longitude. In some cases it might be desirable to use a UTA coinciding with geographical boundaries, and to allow for this UTAs with polygonal outlines may also be used.

1.4 Method of Generalisation

The generalisation method which has been adopted characterizes each UTA by its gray scale frequency distribution. The parameters used for generalisation could be quantities derived from those distributions, such as the mean and the standard deviation, or they could be the histogram that represents that distribution. The latter will be used in this investigation, because the gray scale histograms are the simplest way of handling the information contained in the frequency distribution. For most purposes the four MSS bands will be treated independently, but greater sensitivity could be achieved by using a frequency distribution in four dimensional measurement space.

More complex methods of generalisation such as the power spectrum of the structure in the image could give parameters related to the photogrametric use of texture. The use of such alternative means of generalisation will be investigated at a later stage of the project.

1.5 Data Base For this Investigation

Two test sites have been chosen, although adequate coverage is only available for one of these, which is an area in the Central Valley of California, and includes part of the Sierra Nevada and the Coastal Range. The second test area is in Eastern England.

The 70 mm bulk negatives for each of the scenes including part of the test areas have been examined, and CCTs obtained for 4 of the scenes. The centre points, scene identifiers and local tape code names for these tapes are listed in table 1.

TABLE 1

Local Code	NASA Scene Identifier	Centre Point	Date Taken
TAPE B	1038 - 18114	37° 27'N, 120° 22'W	30 Aug.73
TAPE C	1031 - 10334	53° N, 0°W	23 Aug.72
TAPE D	1056 - 18114	37° 23'N, 120° 22'W	17 Sept.72
TAPE E	1308 - 18122	37° 34'N, 120° 37'W	27 May.73

1.6 Scope of this Report

This is a detailed progress report covering all aspects of the investigation during the period December 1972 (when the first CCTs were obtained) until February 1974. Significant results are presented concerning the accuracy with which the UTAs can be located in CCT coordinates, and the extent to which small scale structures in the gray scale histograms are related to real features on the surface. In addition, progress in the development of analytical techniques is presented.

2. UTA Location and Compilation

2.1 Introduction

Of primary importance for automatic generalisation of satellite imagery, is the ability (i) to align the digital data to geographical coordinates, and (ii) to select those pixels which lie within a specified target area.

There are several ways in which this alignment could be made, but for an automatic system operating on a global scale, the most hopeful in terms of computing time is to use the information available on the satellite position and attitude. This is the method used by NASA for meteorological satellites and for ERTS bulk (system corrected) imagery. From predicted or measured orbital parameters the ERTS processing facility calculates the latitude and longitude of both the sub-satellite point and the central point of the image. The latter corresponds to the intersection of the principal axis of the RBV cameras with the earth's surface and depends on the attitude of the satellite. These two points are then used to superimpose a latitude-longitude grid on the image. The pre-launch estimates of the accuracy of this grid were about 500 m, which implies that a 50km square UTA could be located with a 1% error. This is sufficiently accurate for the sort of changes that could be detected by this method.

2.2 Techniques for compiling the UTAs.

UTAs are defined by the geographical coordinates of their corner points, and are compiled from the CCTs in two stages:

i) The conversion from latitude - longitude coordinates to scan line, picture element coordinates (tape coordinates) is performed by linear interpolation between the tape coordinates corresponding to the

intersection of the latitude-longitude grid with the edge of the frame using the grid information contained in the annotation record at the start of the tape.

ii) The digital data corresponding to a UTA is compiled by selecting from the CCTs those points which fall between lines connecting the corner points of the UTA. The UTA data is then stored as a separate data file consisting of segments of scan line of varying length.

2.3 Results

The accuracy of the coordinate conversions performed by this method have been tested using easily identified control points for the Californian images (Tapes B, D, and E). The geographical coordinates were taken off the USGS 1 : 250, 000 maps, and converted to tape coordinates . A section of the image surrounding the tape coordinate was displayed using line printer gray-scale plots, which gave an indication of the positional errors.

Figure 1 shows the results obtained for the dam of Honey Lake in California. The positions found for the dam on tapes B3 and D3 are marked by open circles, and show errors in the range of 5 to 10 km. Tests for other parts of these images show that the error has a large constant component, and a smaller component which varies across the image and has a magnitude of about 1 km.

2.4 Discussion

The source of the major error is almost entirely in the latitude-longitude grid data provided by NASA, and similar errors are found from the grid superimposed on the photographic products. It is

possible that the minor error is due to the coordinate conversion algorithm. This is particularly likely near the corners of the images where grid marks are sparse, and linear interpolation is less reliable because of geometric distortion.

These errors in the positioning of the digital data relative to the UTAs are too large to allow reliable monitoring for changes in the imagery.

Two solutions are possible for the purposes of this investigation. Either the UTAs can be defined purely in terms of tape coordinates and subsequent images aligned by use of a correction which initially would have to be determined manually, or the geographical coordinates could be converted to tape coordinates with due allowance made for the error in the grid. The advantage of the first is that it removes the need for unpacking the annotation record, and so reduces the core store required for programs, which would bring useful savings in computer usage. The second alternative allows geographically significant (or at least interesting) areas to be used more easily, but at the expense of increased computer usage.

An interesting excercise in this context is the use of a cross-correlation technique to align one image with another. This is a standard technique in crystallographic and in biological image processing where the image consists of well defined objects seen against a uniform background, but it is not immediately obvious that it will be successful in aligning remotely sensed images where there is no clear distinction between object and background. Density slicing is one

means of reducing images to a collection of objects on a uniform background, and this offers the most promising line of action. This approach is currently being pursued.

Alternatively, the tape coordinates of features recognisable on the photographs could be taken directly from a coordinate grid superimposed on the photograph, provided that the tape and photographic imagery are coincident. However, it appears that there is an along-track displacement of tape imagery relative to the photographic imagery by as much as 8km, and so there is no advantage in using this method. The relative displacement of the two types of imagery has been found for all cases for which we have CCTs.

2.5 Development of the Computer System

The following operational programs for handling CCTs and compiling UTAs have been developed.

1. IDREAD :- Reads and unpacks an annotation record in the ERTS CCT Format.
2. ANNRD :- Reads and unpacks an annotation record, including the latitude-longitude grid data.
3. HEADS :- Produces a line printer listing of the relevant information in identification and annotation records.
4. BULK11 :- Constructs the mask for compiling square UTA's with sides parallel to latitude and longitude grid.
5. BULK12 :- Compiles a UTA using the masks generated by either BULK11 or BULK15.
6. BULK15 :- Constructs the mask for compiling polygonal UTA's with a maximum of eight sides.

In addition, routines BULK13 and BULK14 have been written for use as self contained programs for compiling histograms (smoothed and unsmoothed), and for contouring on the graph plotter rectangular areas of the images defined by tape coordinate boundaries.

Apart from these larger segments, several service subprograms are available, including routines for unpacking and packing the digital data into ¼ words as used on the original CCT's, routines for converting the CCT'S to ICL readable format and for producing grey scale pictures on the line printer.

2.6 Display Techniques

In order to locate recognisable ground features in terms of tape coordinates, some form of display of the data on the tapes is needed. Two forms of picture display are available which are to some extent complementary, and both will be maintained.

i) Line printer gray scale pictures simulate the photographic imagery, and thereby provide a rapid display of the data. However, they suffer from the disadvantages of mapping at too small a scale, and of a limited range of gray tones. For speed, no allowance is made for the skew on the imagery, or for the distortions imposed by the use of a standard line printer, although in principle this could be done. Therefore, the resulting map is not geometrically accurate, and is not comparable , in the overlay sense, with conventional maps.

ii) Contour maps of the gray scales drawn by the graph plotter have corrections for the skew and sampling effects, and have an easily variable scale. In addition there is a wider range of gray scale information

available. The disadvantages are the time taken to produce a single display, the difficulty in contouring areas of rapidly changing grey tone and the lack of an immediate visual interpretation of the display.

Figure 1. provides an example of a gray scale printout and Figure 2 shows a sample contour map.

3. Gray Scale Sampling

3.1 Introduction.

Radiometric errors in the CCT imagery impose limitations on the use of the gray scale histograms. It is necessary to investigate the magnitude of these errors, so that the significance of particular differences between histograms can be assessed.

The sources of radiometric error to be investigated are:-

i) those related to the conversion from radiance to gray scale levels on the CCTs

ii) those caused by the uneven response of the sensors.

3.2 The MSS Transfer Function

A detailed description of the MSS transfer function, which determines the conversion from radiance to gray-scale values is given in the section "System Performance" in the ERTS Data Users' Handbook. A brief summary will be given here.

The radiance, R, to tape count (gray scale values on the CCT), T, transfer function is always linear, but as will be seen, the a priori probability of any pixel having a tape count T is a function of T itself: that is, the probability distribution (as opposed to the frequency distribution), is a non-uniform function of T. The cause of this non-uniformity is the discrete nature of the gray scales, and becomes more apparent when the intermediate transfer functions, f_1 between R and the sensor count, S, and f_2 between S and T, are considered.

For band 7 alone, the transfer function f_1 is linear and S has a range of 0 to 63. The second transfer function f_2, which is related to the processing of the raw sensor data by the Special Processing Subsystem (SPS), is also linear and T has the same range as S. Among

the operations performed by SPS is radiometric calibration of the sensor data and this could cause f_2 to differ from a one-to-one mapping of S onto T. If for example two different sensor counts, S_1 and S_2, are always mapped into a single value, T_1, then the *a priori* probability of obtaining T_1 will be twice the mean probability.

The non-uniformity introduced by this effect will be spread over adjacent gray levels only and could be removed by smoothing with a simple running mean technique. In practice the calibration function changes continually, and this smoothing does not appear to be necessary. This is shown by Fig.3 which is the unsmoothed band 7 histogram for a 50 km square area in the central valley in California.

The sensors for bands 4, 5 and 6 are usually operated in the compressed data acquisition mode, in which the sensor transfer function f_1 is non-linear, with the sensor count scale compressed for high radiance values. The range of S is again $0 \leqslant S \leqslant 63$. To correct for the non-linearity of f_1, the SPS transfer function f_2 is also non-linear, and can be approximated by a quadratic of the form

$$T = 0.025\ S^2 + 0.45\ S \qquad (1)$$

This gives a tape count scale which is linear with a range $0 \leqslant T \leqslant 127$, or twice the range of S. Because of this difference in the ranges of S and T, the transfer function f_1 cannot be a one-to-one mapping. Without

the smearing introduced by the changing calibration function which modifies equation (1), the probability of obtaining at least half of the T values would be zero. Because of the non-linearity low T values will correspond to 2 or more S values, whereas at the high end of the radiance scale only one in every two or three T values would have a non zero probability. This is illustrated by Figure 4 which shows the mapping of the 64 S values into T, on the assumption that f_2 is represented exactly by (1). This distribution shows that the a priori probability of obtaining T = 0 is approximately six times greater, on average, than that of obtaining any particular value of T greater than 120.

3.3. The Effect of the non-uniform probability on the histogram.

The grey scale histograms derived from the imagery are the product of this non-uniform probability distribution and the actual grey scale frequency distribution. The effect on the histograms is illustrated by Figure 5. Here each line shows the histogram for 12 scan lines of the band 5 for Tape D3. The tape counts have been combined in pairs to give a range of 64 values. One of the most prominent features of the histograms are the columns with anomolously low frequencies which persist over several sets of 12 scan lines, and which are associated with grey levels for which the a priori probability is low or zero.

Because the exact form of the mapping f_2 from S to T changes from scan line to scan line, and from date to date, the lower limit of the detail in the histograms which can be accepted as real is approximately

the range of T which corresponds to a single value of S. Because of the non-linearity of f_2, this is a function of T, i.e. smaller changes can be detected in the radiance of dark features than can be detected in brighter features.

The irregularity of the histograms can be reduced, and at the same time the redundancy in the grey scale values due to the expansion of the range by f_2, can be removed by smoothing the histograms with a suitable convolution function. The aim of this convolution is to derive a 64 level scale for which each level has an equal <u>a priori</u> probability. The probability of obtaining a value between T and $T + \Delta T$ is related to the probability of obtaining a value between S and $S + \Delta S$ by

$$p(T).\Delta T = p(S).\Delta S \frac{dT}{dS} \qquad (2)$$

For a uniform distribution in S, this reduces to

$$p(T).\Delta T = \frac{1}{64} \cdot \frac{dT}{dS} \qquad (3)$$

Differentiating (1) yields

$$\frac{dT}{dS} = 0{\cdot}05\,S + 0{\cdot}45$$

which, on substituting for S gives

$$\frac{dT}{dS} = 0{\cdot}05\sqrt{81+40T} \qquad (4)$$

The width of the 64 intervals in T which have equal <u>a priori</u> probabilities is found from (3) and (4)

$$\Delta T = 0{\cdot}05\sqrt{81+40T} \qquad (5)$$

The centre points of these intervals is given by

$$T_n = 0 \qquad n=1$$
$$T_n = \tfrac{1}{2}\Delta T_n + \sum_{i=1}^{n-1} \Delta T_i \qquad (6)$$

Preliminary tests have been made using a trapezoidal convolution function which has a width at half height given by (5), outer edges with slopes of +1 and - 1, and an equal weighting for all points near the centre of the interval (Fig.6)

This convolution function has the advantages that it is smoothly varying over the range of T, and allows for the one-to-two or one-to-three mapping at high T values by weighting equally all values near the centre of the window. Figure 7 shows a histogram both unsmoothed and smoothed using this convolution function. The full 128 T values have been used in plotting the histogram, although at most 64 of them are independent, and a 64 point non-linear scale would be more approprate.

3.4 Identification of Scanner Irregularities

The multi-spectral scanner that produces the radiometric data has 6 independent sensors in each band, and thus any unevenness in the response of those sensors will produce a banding effect on the image which would repeat every 6 lines. In an effort to identify the magnitude of this effect, data from the CCTs were subjected to the technique of power spectrum analysis.

This technique of analysis takes a given sequence of data and reduces it to a series of wave forms. Each wave, which is of the form of a sine wave, is characterised by two parameters, a wavelength (the distance between two successive peaks), and an amplitude (the height of the wave form). The power associated with each wavelength is a measure of the contribution of that wavelength to the total signal.

Thus any wavelength corresponding to a pronounced banding effect will plot as a peak on the graph of power against wavelength, which will thus identify strong periodicities present in the data.

The analysis was implemented by use of the program published by Davis (1973). The data input was a string of gray-scale values at right angles to the scan lines, parallel to the flight path of the satellite, with a sampling interval of one scan line. Thus any periodicities with a wavelength of 6 times the sampling interval could be associated with unevenness in the sensors.

The resulting power spectra for all four spectral bands were calculated for several images over lengths of 300 scan lines. The results for an area of the North Sea (Tape C4) and for an area of the Sierra Nevada of California (Tape D3) are shown as Figures 8 and 9. Results, not shown here, were obtained for adjacent scan positions, which demonstrated that the spectra were identical or very similar if a small horizontal shift was applied. The resultant plots have been smoothed by application of the Hanning triangular window (Davis 1973), and power is plotted on a logarithmic scale.

The results for the Wash image (Figure 8), shows that there is a distinct peak near the wavelength of 6 scan lines, identifying the unevenness of the sensors. On this image all the values are very low, reflecting the fact that the image consists of open sea which is uniformly dark. Because of the low input values the resultant spectra are very noisy. However, the peak at wavelength 6 is markedly larger

than the other peaks, which are associated with noise. The height of this peak decreases as the mean radiance increases, implying that the magnitude of the unevenness is similar on all bands, but that its relative importance varies inversely with radiance. The graph suggests that the variability attributable to this effect is of the order of $\frac{1}{2}$ to 1 gray scale value.

The results for the Sierra Nevada (Figure 9) show the power spectra in the presence of a much stronger signal. The parallelism of the spectra for the four bands shows that, for this image, most of the power is related to surface features. Once again a peak is visible at a wavelength of 6, and suggests an irregularity of about 1 gray scale value, that can be attributed to the scanner.

3.5 Discussion

It has been shown that both the non-uniformity of the radiance to gray scale transfer function, and the use of 6 independent sensors for each spectral band, cause detectable effects in the imagery.

The irregularity of the histograms is accounted for by the non-uniform distribution of *a priori* probabilities of gray scale values. This is most significant at high radiance values, where it is equivalent to an uncertainty of about 3 in the gray scale.

The banding introduced by the unequal responses of the 6 sensors causes an uncertainty of about 1 in the gray scale. This is most significant at low radiance values where the errors introduced by the transfer function are small.

These two effects impose a lower limit of 2 gray scale values in the width of the classes which can reliably be used in preparing and comparing gray scale distribution histograms. An algorithm has been suggested which divides the full range of T into 64 classes with equal *a priori* probability, but the importance of sensor unevenness at low radiance levels implies that 32 classes would be more suitable.

4. The Analysis of Gray-Scale Histograms

4.1 Introduction

Initial examination of histograms for several areas showed that there was a possibility of equating components of the histogram with surface types. Thus the possibility of splitting histograms, and thereby deriving estimates of the percentage of any image covered by each surface type was investigated.

The investigation concentrated initially on an area which showed a simple bimodal histogram, which could be readily interpreted. This area was part of image C4, which covered part of eastern England and the adjacent sea area. The land in this area consists mainly of small fields of the same order of magnitude, or smaller, than the 70 m resolution of the scanner, and thus the overall impression given by the land is of a uniform gray speckle. It was thus not expected that this technique would be able to distinguish between different areas on the land. In contrast, the sea on the image shows up as a dark area, visible on all four bands, although the greatest contrast is offered by band 7, in which the gray scale values for the sea are nearly all either 0, 1, or 2. However, on bands 4 and 5 the considerable areas of mud-flat, turbid water and sand bank are readily visible. It was hoped that for these bands it might be possible to split the histogram into three components, thereby identifying proportions not only of land and sea but also of turbid water. However, the histograms were in fact monomodal showing that the two distributions that could be identified on bands 6 and 7 were overlapping, and that it was not possible to accurately identify even two components of the histogram.

4.2 Technique adopted

A large number of techniques have been developed for splitting an observed frequency distribution into the sum of two or more component distributions. Clark (pers.comm.) has reviewed these methods, and recognised three main groups; graphical, analytical, and numerical.

Graphical techniques rely on the subjective interpretation of discontinuities in the cumulative frequency curve, and are thus subject to a high degree of operator error. Analytical techniques are exceptionally difficult to calculate, and do not on every occasion yield a solution. Numerical techniques are much more varied, but nearly all use iterative methods to improve upon initial estimates of parameters describing the component populations. They are thus suitable for inclusion in an automated data handling system.

Of the several alternatives available, it was decided to adopt that of Jones and James (1972), which uses a maximum likelihood method of estimation. The main reason for this decision were as follows:

a) Although initial estimates of the parameters to be estimated are required, the method will converge to a stable solution even if these are considerably in error, and it would thus always be possible to provide a complete null set of initial estimates.

b) The published work included a FORTRAN program, and so made it possible to examine the capabilities of this approach without the investment of a large programming effort at this exploratory stage.

c) It is possible to extend the existing program to include types of component distribution other than the Normal. The published work includes both the circular Normal distribution, and the mechanism for adding other distributions. One major disadvantage is that at present the existing program will handle only two components.

4.3. Brief Outline of the Maximum Likelihood Method

The maximum likelihood method relies on the fact that for any observation set, X, and a statistical model with parameter set θ, it is possible to calculate the a posteriori probability of X occurring as the outcome of the stochastic process defined by the model. Thus for any model it is possible to find, in principle, a parameter set θ that will maximise the a posteriori probability, or likelihood, L, of the observation set X.

In the case of a set of parameters refering to a mixture of distributions, the model becomes:

Let the j^{th} underlying probability distribution be $f_j(x_i)$, where $j = 1, \ldots, m$, and the subscript i, $i = 1, \ldots, n$ refers to the observations. Thus for each component distribution, the probability of the observation set is:

$$P_j(X) = \prod_{i=1}^{n} f_j(x_i)$$

If the proportion of each component distribution is Q_j, then the total a posteriori probability of the observation set is given by

$$L(X) = \sum_{j=1}^{m} Q_j \prod_{i=1}^{n} f_j(x_i)$$

Usually, however, it is easier to transform these probabilities into their logarithms, and replace the multiplication by an addition, so that the transformed expression to be maximized becomes

$$\lambda(x) = \sum_{j=1}^{m} Q_j \sum_{i=1}^{n} \log f_j(x_i)$$

The parameter set θ for the model consists of (m-1) values of Q_j, and the parameters for each of the sub populations f_j. Note that because the Q_j's must sum to unity, (m - 1) values fully determine set.

In the case to be considered here, the underlying distributions are all considered to be Normal, and thus have the form

$$f(x_i) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-(x_i-\mu)^2 / 2\sigma^2\right\}$$

where μ and σ are the mean and standard deviation. Where there are two such distributions, only a single proportionality parameter is required, which we will call α.

Thus the total statistical model to be considered is

$$f(x) = \alpha\left[\frac{1}{\sqrt{2\pi}\,\sigma_1} \exp\left\{-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right\}\right] + (1-\alpha)\left[\frac{1}{\sqrt{2\pi}\,\sigma_2} \exp\left\{-\frac{(x-\mu_2)^2}{2\sigma_2^2}\right\}\right]$$

And the parameter set thus consists of the five parameters

$$\theta = \{\alpha, \mu_1, \sigma_1, \mu_2, \sigma_2\}$$

Because the derivatives are non-linear, it is not possible to reach the maximum likelihood solution by direct evaluation, but it is necessary to use a numerical technique. The Jones & James Program uses the steepest ascent method to provide a rapid convergence towards the maximum, and thereafter changes to the Newton Raphson method which is superior near that maximum, but also more sensitive to poor initial estimates.

4.4 Application to the Problem

In order to be able to use the maximum likelihood method it is necessary to have an observation set, rather than a grouped histogram. However, the histograms of any one area are the result of the examination of a large number of pixels, and it was thus considered convenient to convert the histogram to a reduced observation set, purely as a form of data reduction. Because of the integer nature of the gray-scale values, it was possible to reconstruct an observation set with approximately the same proportion of values at each gray scale level as in the original data set. The program was set up to accept a data set of up to 200 points, which was constructed from histograms typically derived from 80,000 points. Repeated analysis using a larger data set of 400 points yielded almost identical component distributions, confirming that this data reduction does not distort the results.

The efficiency of the program was improved by the use of integer, rather than real, arrays for storing the data set. Although this required the use of the FLOAT operation for calculations of the probability function, it permits a more efficient storing of the data. A histogram drawing routine was provided, which used the line printer to give a plot of the input histogram, the histogram derived from the initial estimates, and the histogram fitted by the program.

4.5 Results

Results were obtained from the program for histograms derived from all 4 spectral bands. The data from band 7 yielded the clearest results, since the bimodality of that histogram was extreme, with virtually no overlap between the two components. Initial estimates of the parameters were chosen visually, and a total of 30 iterations were required to produce stable estimates. A sample output is included as figure 10, and a summary of the results for band 6 and 7 is given below:

Results of Maximum Likelihood Estimation of Parameters of Mixed Normal Distributions

Parameter	Band 7		Band 6	
	Initial	Final	Initial	Final
Proportion of pop^{n} 1	.7	0.717	0.7	0.769
mean of " 1	1.0	1.255	3.0	2.856
s.d. of " 1	1.0	0.712	.4	1.478
mean of " 2	23.0	22.829	20.0	20.562
s.d. of " 2	5.0	5.576	3.0	3.36

As can be seen from the table, for band 7 population 1, which was identified as the sea component, covers 71.7% of the area. The corresponding estimate for band 6 however, was higher, at 76.9%.

The results for bands 4 and 5 were considerably less encouraging. The likelihood response surface proved to be very flat, and thus no stable solution seemed possible. In particular, the estimates of the standard deviations seemed most unstable.

Furthermore the most stable parameter, the proportion of the data in population 1 converged towards a value greater than one. In a different context such a result, and the corresponding negative proportion for population 2 can be interpreted as representing a single parent population from which a distinctive sub-population has been removed. However, in this case the concept of a negative (subtractive) population has no interpretable meaning, and the results are thus of little value.

4.6 Discussion

This pilot study has shown that the automated splitting of gray-scale histograms is possible, and can yield useful results. It has also shown that the results yield discrepancies between bands which were sufficient to cast doubt upon the applicability of the method in general. In the light of these considerations, and the considerable computational time required to perform the analysis, it was felt that this approach, while possibly useful in specific instances, was not suitable for routine application to ERTS data.

The alternative histogram splitting techniques that are available might produce results more rapidly, and be more readily applied. In particular methods that can handle more than two components need to be examined. However, at this exploratory stage it is necessary to make considerable checks on the input histograms, to correlate these with ground truth.

At the time of this part of the study, the problem of correlating tape coordinates with geographical areas, had not been solved. Thus the identification of the spectral responses of land types necessary

for the interpretation of the histograms could not be pursued. This technique was not therefore tested further until more background information became available.

It is however pertinent to consider the question whether other analyses might not yield similar results as cheaply. In particular if it were possible to map ground types, it would be a simple matter to estimate their area. The possibility of using either cluster analysis to group points (or collections of points) into hopefully recognizable groups, or some form of discriminant function to map points onto a priori groups might be equally useful. Some of these techniques are discussed below.

5. The Use of Clustering Techniques

5.1 Introduction

In an attempt to produce a generalised map of the ERTS digital data the possibility of using cluster analysis was investigated. This form of analysis would group like areas, and hopefully each cluster would be identified with a land type. Clustering using the spectral signature of individual points has been used by many workers (eg. LARS, 1968).

Because of the very large amounts of information available from one ERTS CCT, it is useful to set up some intermediary units that represent an aggregated data points, which become the Operational Taxonomic Units (OTUs) which are the input to the clustering procedure. Once they have been designated, it becomes possible to collect data for these units from the digital tapes. These units are then input into a clustering algorithm which produces as an output a map of land types for a large area. There are thus three basic steps in the analysis.

a) Setting up the OTUs.

b) Selecting the variables to be used.

c) Producing the clusters.

5.2 Setting up the OTUs

The size of the OTUs considered has the effect of setting a scale for the resultant map, whereas their number determines the computational effort required.

Setting aside, for a while, the computational problems, several effects can be noted which are related to the size of the OTUs:

1) A lower bound for the size of the OTU is set when each OTU is a single pixel. At this scale, it is probable that the 'noise' present in the picture would be sufficient to mask any regional patterns that might be detected.

2) As the size of the OTU increases, so does the amount of information available about its characteristics. The complete information is given by the full set of gray scale values, spatially ordered, and this information amount obviously becomes very large as the OTU increases in size. This full data set contains much information that is not relevant at a regional scale, and it is thus necessary to derive a variable set which effectively characterises that information. Obviously the larger the OTU the easier it is to derive an efficient variable set. For example attempts to measure texture require that each OTU contains a fairly large number of pixels.

3) As the size of the OTUs increase, so the resolution of the final map diminishes. Thus boundaries between land types can be more accurately, but more expensively mapped if the OTUs are small, but on the other hand larger OTUs permit a classification on a much larger areal scale.

4) As the size of the OTUs decrease, so their number increases and so the final classification may alter. In particular, small OTUs may, by chance, contain anomalous pixels, which would result in some OTUs remaining outside nearly all the clusters until very late in the clustering procedure. Large OTUs would tend to reduce the number of these 'Maverick'

points, since they would be included in, and hopefully 'swamped out' by larger numbers of normal pixels.

With these considerations in mind, the problem of classifying a significant portion of an ERTS image was considered. Initially the target area to be clustered was set at $\frac{1}{16}$ of an image, that is the data contained on $\frac{1}{4}$ of one CCT. This represents 810 x 500 pixels, which is about 50km x 50km. For the purposes of initial investigation OTUs were arbitrarily set a size 20 x 25 pixels. This size of OTU has the advantages that:-

a) Each OTU contains 500 pixels, which is sufficient to permit the evaluation of all parameters that might be considered. In particular it is sufficient to give relatively smooth gray-scale histograms (see section 3, above).

b) The target image thus consisted of a matrix of 40 x 20 OTUs which was felt to be sufficient to produce a fairly detailed map. However, the 800 OTUs thus produced are more than it is feasible to use in a clustering algorithm which requires the entire interdistance matrix, since this would require the storing and repeated reaccessing of 640,000 distances. Alternatively by the use of algorithms that require only portions of the interdistance matrix, it will be possible to cluster these 800 points. However, the consideration of a smaller portion of the image as a training set is also feasible.

c) A subsidiary advantage of these rectangular OTUs is that they can be plotted on a computer line printer without geometric distortion if so required. It must be remembered that this size

of OTU is, however, essentially arbitrary, and was adopted purely to make OTUs available for an initial investigation. A full examination of the effects of adopting other sizes of OTU must be undertaken. However, it is obvious that the choice of OTU size can be made only in relation to an *a priori* objective. This choice, however, requires background knowledge of the effect of changing OTU size.

5.3 The Variable Set.

With an OTU containing 500 pixels, each of which contains a gray-scale value in 4 spectral bands, some sort of summarising procedure is necessary. Several possibilities seem available.

i) Summary statistics traditionally used to describe frequency distributions could be used.

The measures that seem most suitable would be the mean, the standard deviation, and perhaps the skewness and the kurtosis. Taking the mean and standard deviation would reduce the 500 pixels to an 8 variable set, and thus achieve a great compression and eceonomy. However because of this very great compression it was decided not to use this variable set in the first instance, but to use a more comprehensive variable set, and then to determine if such a reduced set would yield the same results. If it turned out that this reduced set would yield satisfactory results, its

efficiency would certainly lead to its adoption. The use of this reduced set will be exained in detail at a later stage.

ii) the variable set adopted for initial consideration consisted of the information contained in the gray-scale histograms.

The gray scales for the 4 bands were all split into a series of classes, and the number of observations within each of these classes computed. The number of pixels within each of these gray scale classes was thus taken as the variable set for input to the clustering algorithm. A first study split the gray scales to a 32 variable set, with 8 equal classes for each spectral band.

For further investigations, the gray scale will be split into 16 classes per band, thereby producing a 64 variable set. It is hoped that this very large number of variables will contain nearly all the information available at this scale, and will contain considerable redundancy. It is then planned to examine the redundancy present, in an attempt to produce a reduced variable set. This may be achieved by use of a stepwise discriminant analysis, which will arrange the variables in their order of importance in determining the groupings. Other variable sets will also be compared with results from this full set. In this way it is hoped to be able to use this extremely full variable set to act as a yardstick against

which to measure the performance of smaller variable sets.

iii) Alternative variable sets can be devised which rely on more sophisticated transforms of the data.

For instance, the ratios of signals in various bands have been found useful by some workers for specific instances, and this kind of variable might prove valuable in clustering for those special purposes. A second instance that is being considered is the use of a Fourier transform to measure texture, and then to input the coefficients of the 2 dimensional Fourier transform in an attempt to produce clusters based on textural properties. Other data transforms such as the Hadamard and Walsh transform (Harmuth, 1969) might also be used.

A last possibility that might be examined is the use of a technique such as principal components to produce a reduced variable set. This would require the examination of the entire image to produce the components and then to cluster on a reduced set of component scores

5.4 The Clustering Program

The data for the OTUs are then input into a clustering algorithm which produces groups of points in the variable space,

which hopefully can be identified with land types. All available clustering algorithms use some measure of similarity (or dissimilarity) between the OTUs in the variable space, and most require that the entire interdistance matrix be stored. The clustering algorithm then selects the two most similar points in the variable space (which is usually assumed to be Euclidean). These two points are then replaced by a single point, usually at the weighted mean of the two, and the distance to all remaining points recalculated. The processes is then repeated on the reduced matrix of interdistances, until only a single group remains. This algorithm is due, in essence, to Ward (1964), and many programs have been published which offer varieties of this basic algorithm.

For the purposes of initial experimentation the program published by Veldman (1967) was used, since this was readily available, and had already been used in a similar situation by Owen Jones and Custance (1974). However it is suspected that more efficient algorithms are available, particularly those that require that only one half of the symmetric similarity matrix be stored within the machine. The possibilities of using other programs will thus have to be explored if this technique is finally adopted for use.

5.5 Preliminary Results

A preliminary study used a block of 5 x 5 OTUs, located on the eastern edge of image D3 in the Sierra Nevada of California. The block of observations ran from positions 1 to 100 along the scan line, and used scan lines 126 to 250. The data were expressed in a 32 variable space, so that each of the equal classes represented 16 gray scale values in bands 4 to 6, and 8 gray scale values in band 7.

These 25 points were clustered using the Veldman program, and a dendogram drawn from the results (fig.11). The maps derived from the last 4 steps are shown as fig.12, and the loss of information due to clustering shown as a function of the number of clusters in fig.13.

There seems to be some justification for thinking there is a distinct jump in the error loss function after the 21st iteration, ie. with 4 remaining groups. At this level the map of the resultant groups is beginning to show coherent spatial units, which retain their general pattern through the subsequent clusterings.

Figure 12 shows that the clusters form spatially compact units, and produce a thoroughly credible pattern. The roughly diagonal arrangement is parallel to the prominent ridge and valley structure of the Sierra Nevada.

Although these initial results seem reasonable, this test is very preliminary, and must be supplemented by much more detailed examination of the clusters produced from a larger data set, and then checked against ground truth.

It is however very encouraging that the clusters are spatially coherent, even though their spatial properties are not included in the variable set. This is in contrast with the results of Owen-Jones and Custance (1974) who found that without applying the contiguity constraint, the resultant clusters were not coherent.

5.6 Discussion

Results so far show that clustering of aggregated units using the gray scale histogram is a feasible procedure, and that coherent groups can be produced without the imposition of the contiguity constraint. However the quality of the results must be tested against a larger data set.

The computational aspects of the application, however, indicate that the program will only be useful for clustering small numbers of OTUs. By storing the entire interdistance matrix, the program requires sufficient core to hold an N^2 matrix to cluster N points, which must then be scanned N times. However, it is possible to use clustering techniques that use considerably smaller amounts of core storage. For example the single-link method has been programmed by Sibson (1973) in such a way as to require of the order of 4.N words of store, and a contiguity constrained algorithm has been presented by Openshaw (1974) which requires similarly small amounts of store. Both these algorithms are currently being investigated with a view to their implementation.

6. Edge Detection

6.1 Introduction

Of the several techniques available for the computer processing of pictures, one of the most fundamental approaches is to split the picture into non-overlapping zones by the detection of edges between approximately uniform areas (Rosenfeld, 1969). There have thus been a considerable number of techniques developed for detecting edges in pictures.

It was decided to apply some of these techniques to the ERTS Imagery, in an attempt to identify the kinds of feature that could be distinguished by their sharp edges. In particular it was felt that this technique would be useful for identifying areas on the tape where readily identifiable features, such as lakes, are lacking. The technique would thus be potentially very useful in the irrigated zone of the Central Valley of California, where large spatial units occur in a regular rectangular pattern.

6.2 Techniques

Commonly, edge detection procedures consist of two steps, the first an initial detection of potential edges, and the second a process designed to 'clean up' the resultant image. Such a scheme will be adopted here, though at present only the first step has been implemented. The standard technique in finding edges is to locate the places where there is an abrupt change in the picture function (i.e. in the gray scale values). This is performed by examining the local slope of the picture function, and where this slope exceeds a given threshold value, an 'edge' is detected. Various more sophisticated filters have been

proposed (eg. by Rosenfeld 1970), which examine the differences between average values either side of the potential edge, or which apply matched filters to each point in turn. However, for the purposes of this investigation it was felt initially that a fairly crude definition of an edge would suffice, and that any cleaning up of the picture that this would require could be carried out subsequently.

The operational procedure adopted was to examine each point in turn, and where the gradient at this point exceeded a threshold value, to demark that point as belonging to an edge. Thus an initial program was written which calculated the maximum slope at each point, using the function:

$$D_{ij} = \text{Max}\{ |a_{i+k,\,j+l} - a_{ij}| \} \qquad \begin{matrix} k = -1, 0, 1 \\ l = -1, 0, 1 \end{matrix}$$

Where a_{ij} is the gray scale value at point ij. This value D_{ij} is thus the maximum difference between each point and its eight adjacent neighbours.

6.3 Results

A program was written to calculate these values of D_{ij}, and produce a binary image of edge points. This has been applied to a portion of the central valley of California (Tape D3). Fig 14 shows a gray scale print out of a portion of the analysed area, and Fig.15 the edges detected in that area.

As can be seen, this operation results in edges which closely correspond with the field pattern as revealed in the gray scale printout but has the unfortunate property that all edges are double, because the programme has identified each edge twice. This effect can be explained

by reference to the one dimensional case. If we scan along a line, and find a large difference between points i and i + 1, point i will be labelled an edge. However this same difference will also be located when examining point i + 1, so that both points i and i + 1 are labelled as belonging to an edge.

To circumvent this problem, differences were calculated only for one quadrant of directions, so that only three adjacent points were examined. Thus the edge identifying function becomes

$$D_{ij} = \text{Max}\left\{(|a_{i-1,j} - a_{ij}|), (|a_{i,j-1} - a_{ij}|), (|a_{i-1,j-1} - a_{ij}|)\right\}$$

Two thresholds were used for the plotting of D_{ij}. Where D_{ij} was greater than 6, a dot was plotted, and where greater than 12 an asterisk was plotted. A result of this program, for the same area as for the other two pictures is given in fig.16. Note that a much cleaner image results, and that very few edges are actually lost.

6.4 Discussion

Obviously this work is still at a very preliminary stage, and much more needs to be done. However, the results of these first attempts are sufficiently promising to indicate that this line of approach is valuable. The following tasks are seen as important.

a) the investigation of the use of different thresholds.

b) the drawing of these maps using all four spectral bands, both as four separate maps, and as one single, multispectral map.

c) the development of a routine which will 'clean up' the resultant binary image. This could be done by removing all points not surrounded by at least n adjacent 'black' elements, where n is some threshold value. Work by Yamada and Farango (1965) suggests that a value of n = 4 would be sufficient, but examination of the

concept of a continuous linear edge suggests n = 3 would be more appropriate. Obviously, however, the value of the threshold used here would not be independent of that used to detect edges in the initial picture processing.

d) The extension of the technique to areas away from the regular field systems of Central California. It is hoped that it might be possible to use this technique to locate ridge and valley lines in areas of the Sierra Nevada.

7. Conclusions

Significant progress has been made in the handling of ERTS CCT data to the extent that it is now possible to compile subsets of that data suitable for generalisation, and to perform some preliminary analyses. Results have been obtained from the following procedures.

1) The compilation of data sets for both polygonal and rectangular U.T.A.'s has shown that the NASA supplied latitude-longitude grid is insufficiently accurate to permit direct compilation of U.T.A.'s defined by geographical coordinates. An intermediate step to correct for this inaccuracy is necessary. This step is performed by locating on the CCT's points whose geographical coordinates are known and determining the correction manually.

2) Generalization of U.T.A. data to gray scale frequency distributions has been examined. Studies of the transfer functions and sensor inequalities have shown that differences of 1 or 2 gray scale values are not significant for our purposes, and that the gray scales can therefore only be divided into a maximum of 32 classes.

3) Analysis of individual histograms have successfully identified the proportions of land and sea in a given image. This was successful for MSS bands 6 and 7, but results for bands 4 and 5 were inconclusive and reflect the influence of turbid water in the image.

4) Preliminary tests have shown that satisfactory classification of areas into subsets representing surface types can be made using an

unsupervised clustering algorithm in which the gray scale histograms make up the variable set. Detailed tests of the relevance of the classification have yet to be made.

5) An edge detection routine has been developed which has been used to locate field boundaries and lake edges in the California test site.

The work to date has solved the initial problems relating to selecting areas and generalising data for the UTA's. We are now able to evaluate the effectiveness of various types of generalisation, with the knowledge that differences between the generalised data represent real effects rather than errors introduced by uncertainties in the position or in the gray scale distribution.

In the next reporting period it is proposed to concentrate on the clustering approach, using various size OTU's and different levels of generalisation. Preliminary tests of other methods of generalisation (for example, textural measurments) will be made in this period.

Reference

Davis, J.C. (1973) Statistics and Data Analysis in Geology
John Wiley & Sons. New York.

Harmuth, H. F. (1969) Application of Walsh Functions in Communications.
IEEE Spectrum . vol 6. p82-91

Jones, T. A. & James, W.R. (1972) Maxlike: FORTRAN IV Program for
Maximum Likelihood Estimation. Geocom.Programs. 5.

L.A.R.S. (1968) Remote Multispectral Sensing in Agriculture.
Research Bulletin. no 844. Purdue University

Openshaw S. (1974) A regionalisation program for large data sets.
For Publication in Computer Applications (University of Nottingham).

Owen-Jones, E.S. & Custance, N.D.E. (1974) Digitised Analysis of Skylark
Rocket Imagery Journal of the British Interplanetary Society.
vol 27 No.1. p 18-22

Rosenfeld, A. (1969) Picture Processing by Computer Academic Press, New York.

Rosenfeld, A. (1970) A non-linear edge detection technique. Proceedings.
IEEE vol.58. p 814-816.

Sibson, R. (1973) SLINK: An optimally efficient algorithm for the single-
link cluster method. Computer Journal. vol.16 no.1 p30-34

Veldman, D.J. (1967) Fortran Programming for the Behavioural Sciences
Holt, Reinhart & Winston, New York.

Ward, J.H. (1963) Hierarchical Grouping to Optimize on Objective Function
Journal. American Statistical Association. vol.58 p.236-44

Yamada, S & Farango, J.P. (1965) Experimental Results for local filtering of
digitized pictures. Report 184. Dept. Computer Sciences. University
of Illinois.

FIGURE CAPTIONS

Figure 1. Position errors for the dam wall of Honey Lake.
The open circles show the position of the dam (marked by the hexagon) as found from the latitude - longitude grid included with the CCT annotation data.

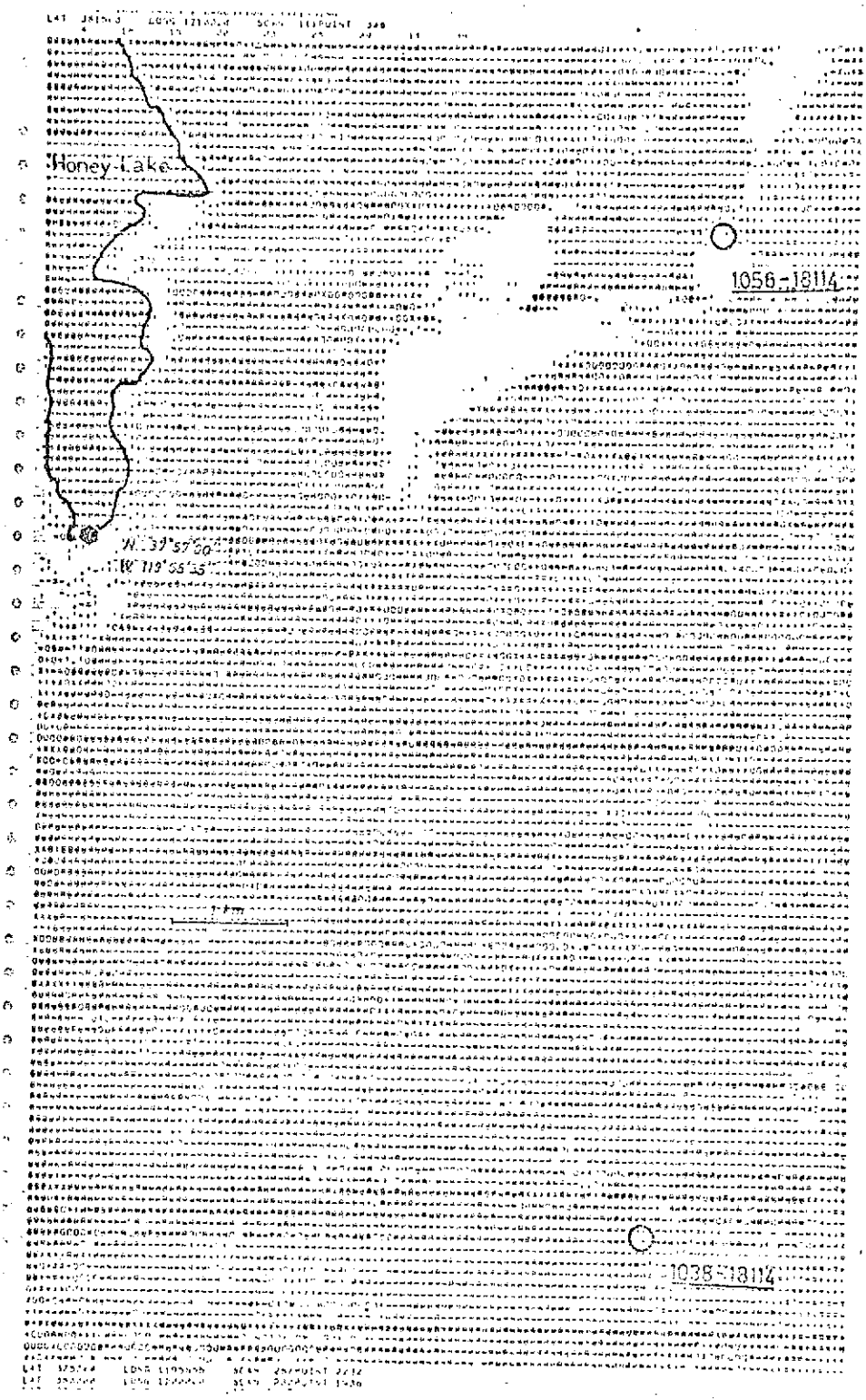

Honey Lake
1056-18114
1038-18114
W 119°55'55"

Figure 2. This contour map of gray scale values for the sea area near the Wash, U.K, gives a good indication of the distribution of sediment laden water.

ERTS IMAGE CONTOURS

IMAGE ID 1081-10314
SCAN LINES 368 TO 1111
MSS BAND 5

CONTOURS AT 10.8 11.8 12.8 13.8 14.8
15.8 16.8 17.8 18.8 19.8

CONTOURS for the WASH

Figure 3. Unsmoothed gray scale histogram for MSS band 7 taken from a UTA in the Central Valley, California. Note how much smoother this distribution is than the corresponding distribution for band 5 shown in Fig.7.

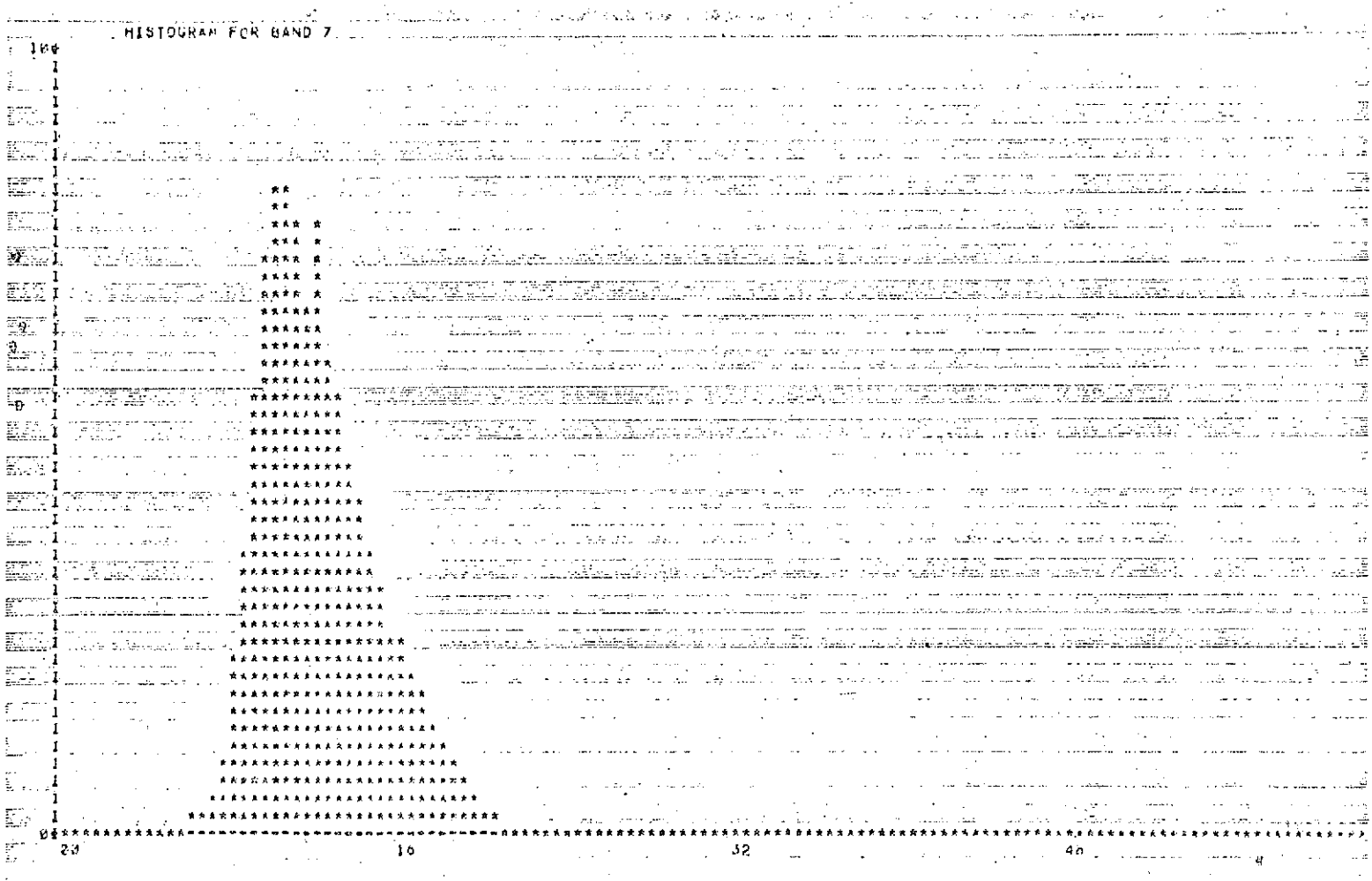

HISTOGRAM FOR BAND 7
100
0
20
16
32
48
4

Figure 4. The mapping of the 64 sensor count values onto the 128 tape count values using the function described in section 3. The vertical axis gives the number of sensor counts mapped into each tape count.

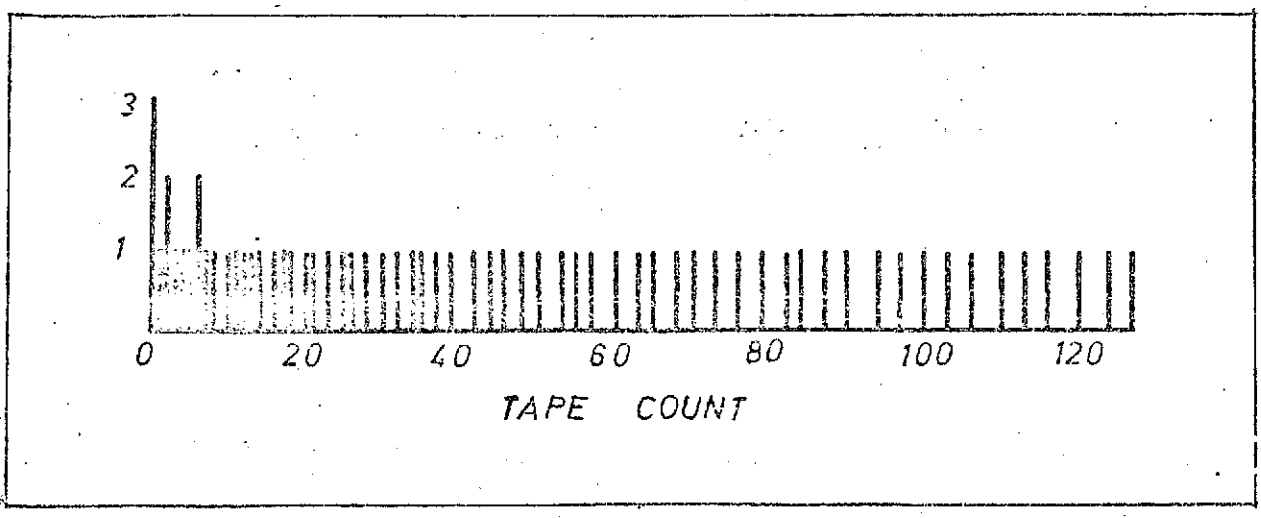
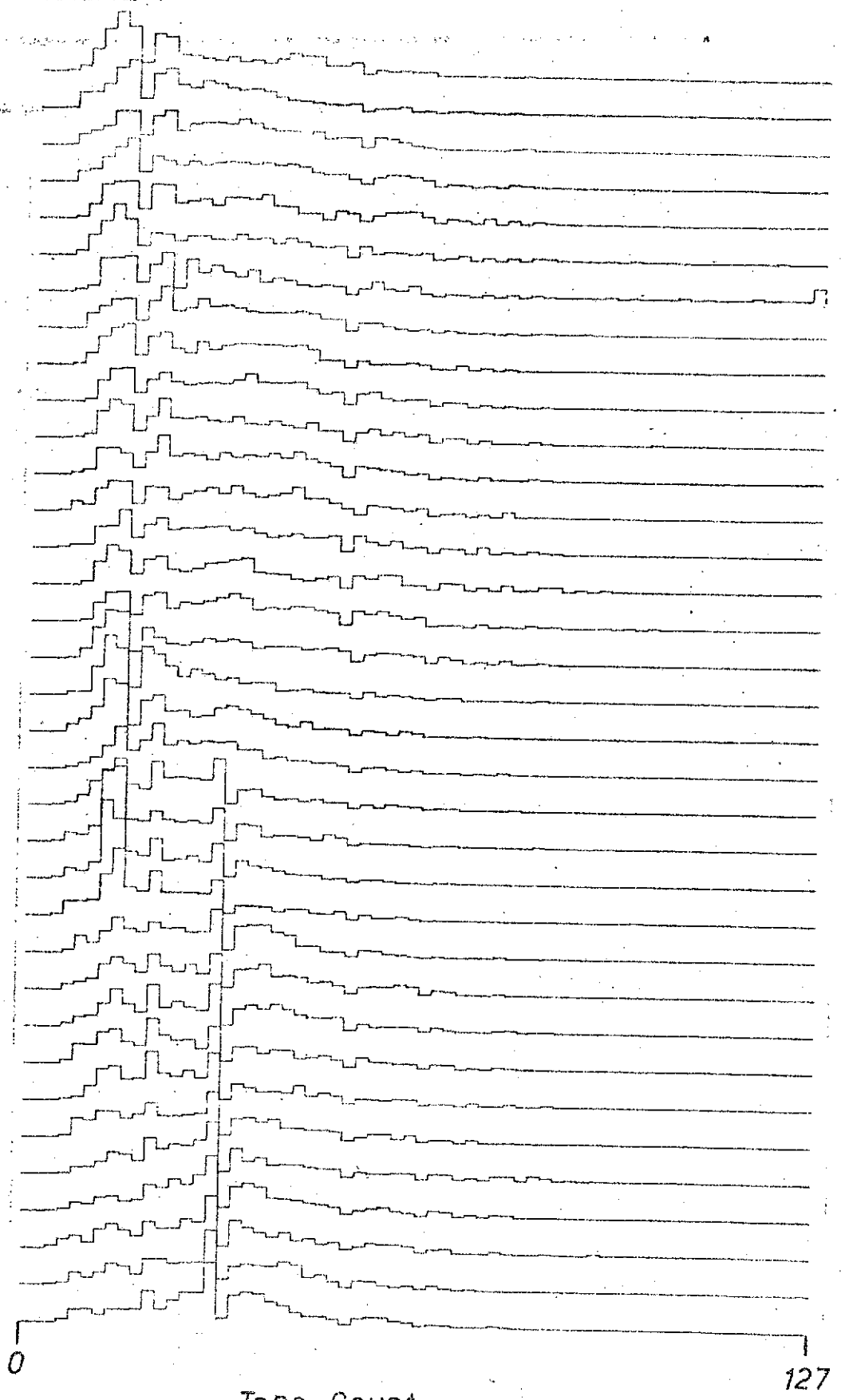

3
2
1
0
20
40
60
80
100
120
TAPE COUNT

Figure 5. Gray scale histograms for successive groups of twelve scan lines for an area in the Sierra Nevada mountains.



Tape Count

Each class corresponds to 2 gray levels.

Figure 6. The convolution function used to remove the effects of uneven tape count probability distribution. The width of the function at half height is given by the equation for ΔT in section 3.

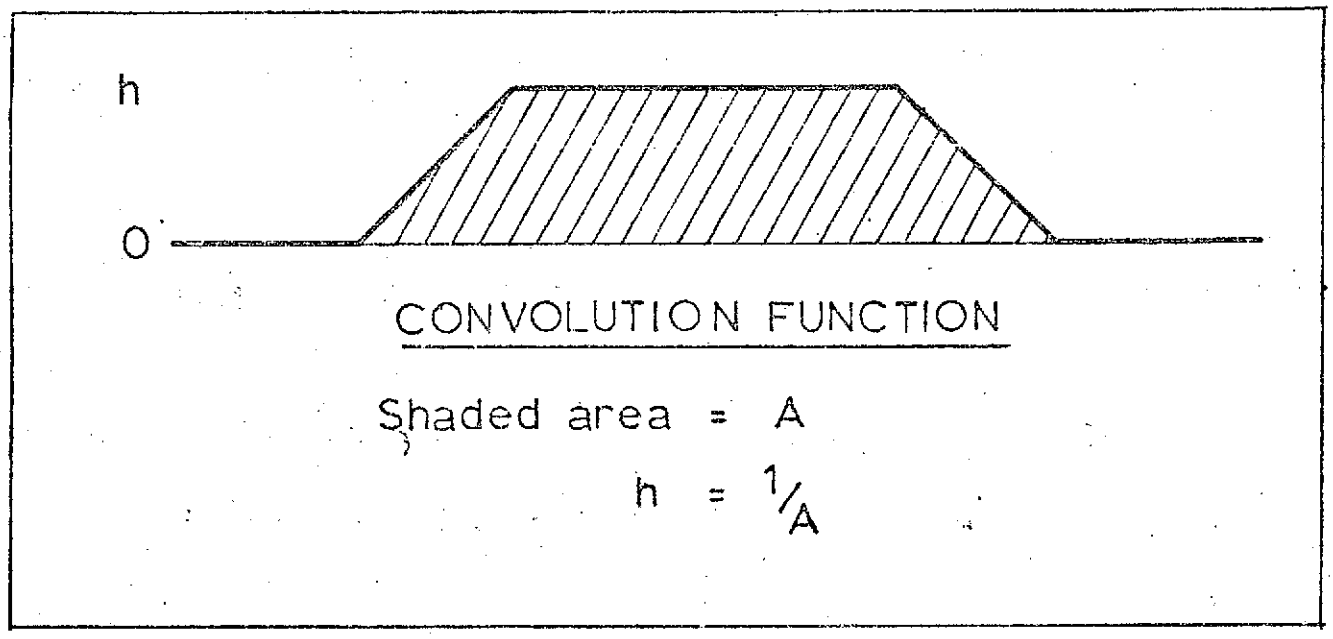
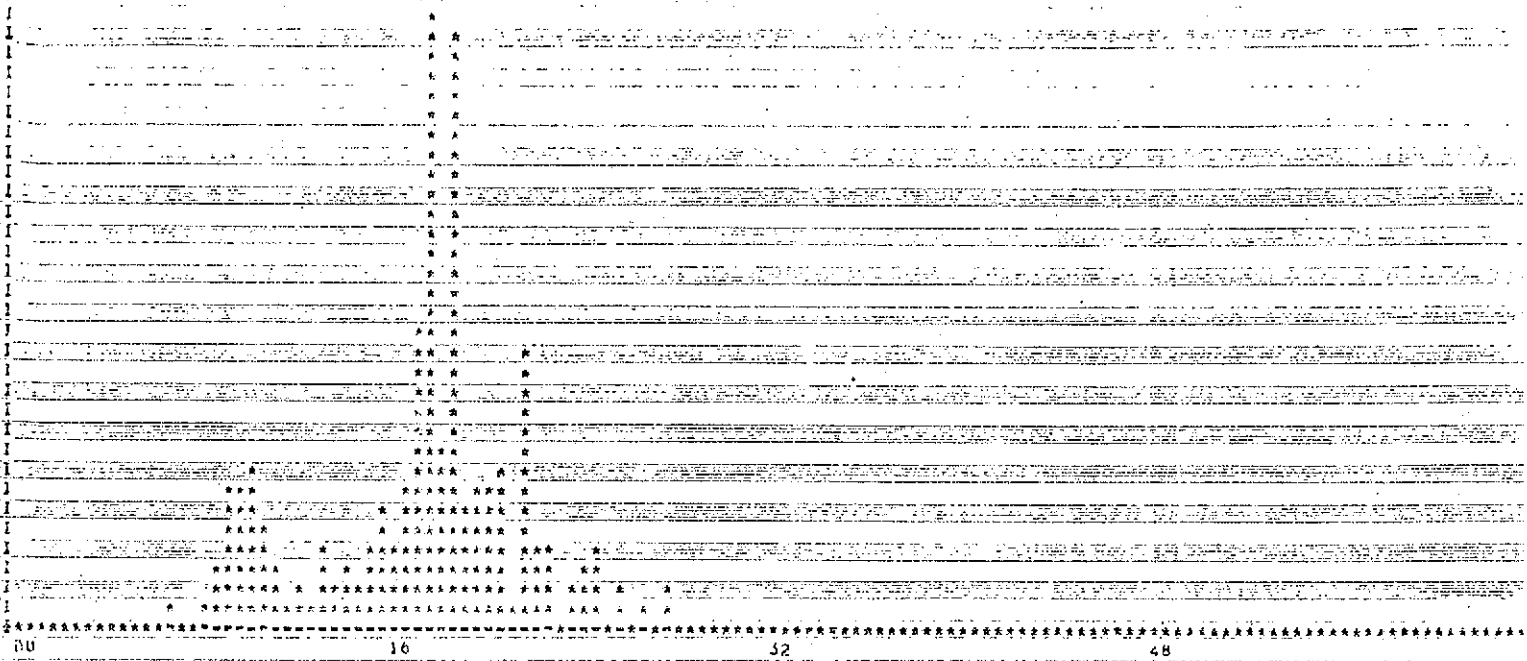

h
0
CONVOLUTION FUNCTION
Shaded area = A
h = 1/A

Figure 7. Smoothed and unsmoothed gray scale histograms for a UTA in the Central Valley.

HISTOGRAM FOR BAND 5



HISTOGRAM FOR BAND 5

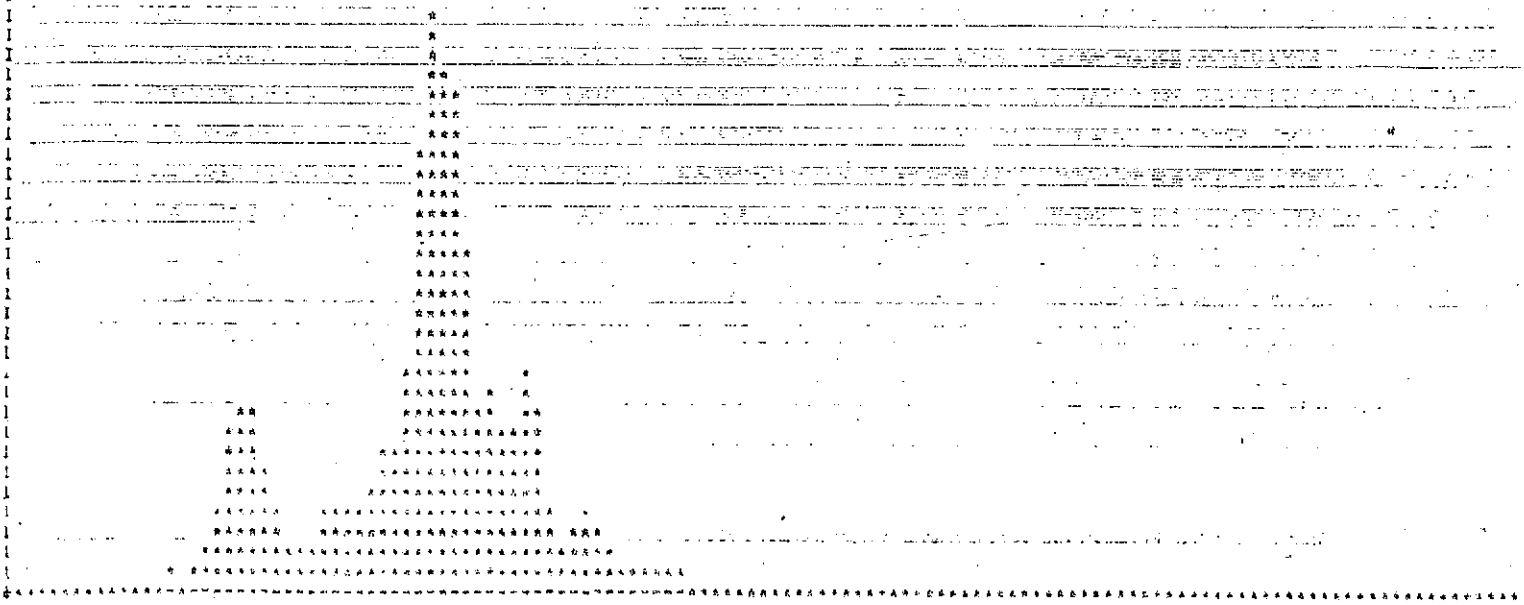


Figure 8. The power spectra for a strip one pixel wide and parallel to the edge of the frame for the sea area included in the U.K. CCT image. The spectra are plotted as a function of wavelength in scan lines, and the power is plotted as dB relative to the noise power in band 4.

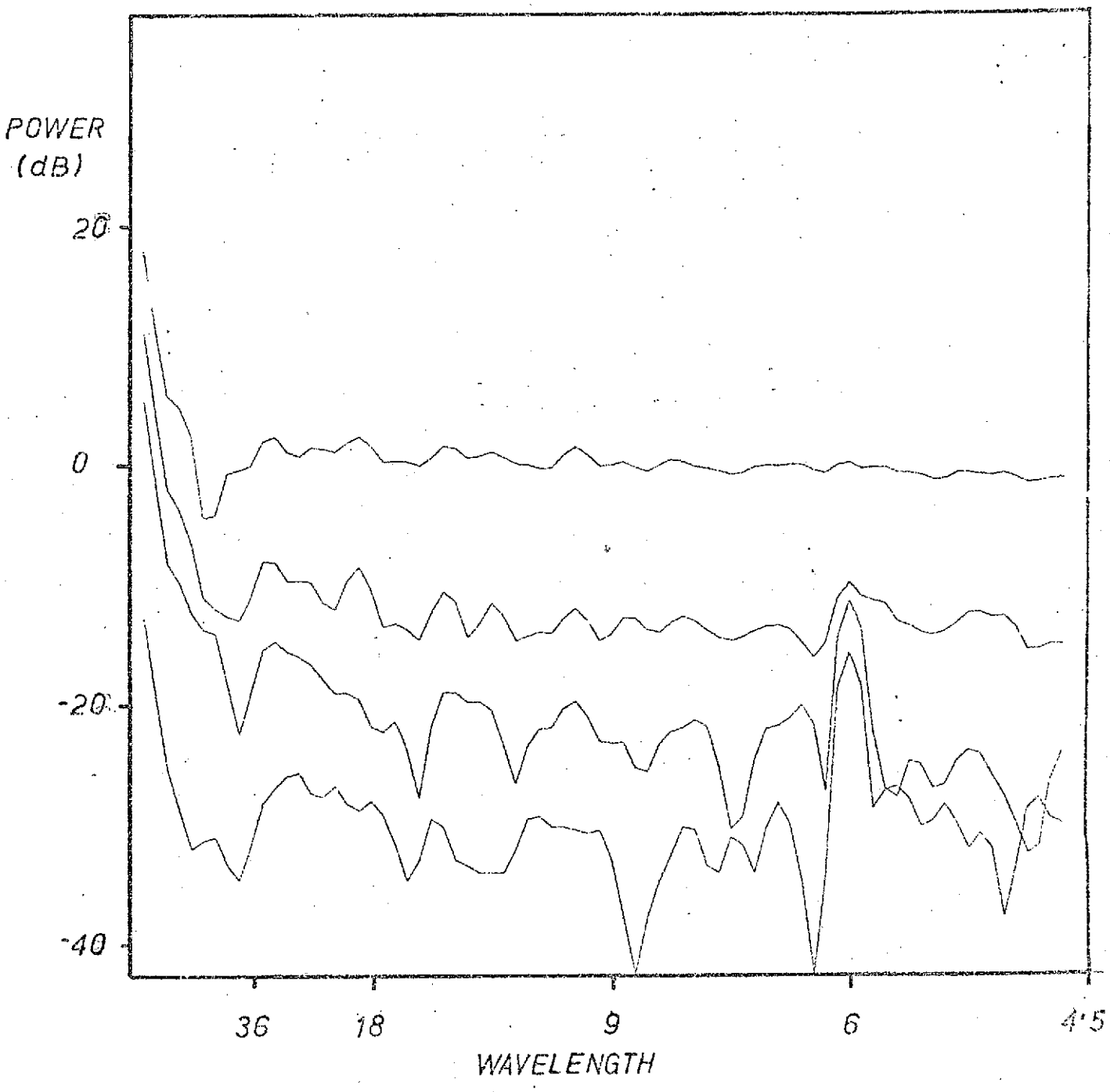

POWER
(dB)
20
0
-20
-40
36
18
9
6
4·5
WAVELENGTH

Figure 9. The power spectra for a strip through the Sierra Nevada range. Axes are as for Figure 8.

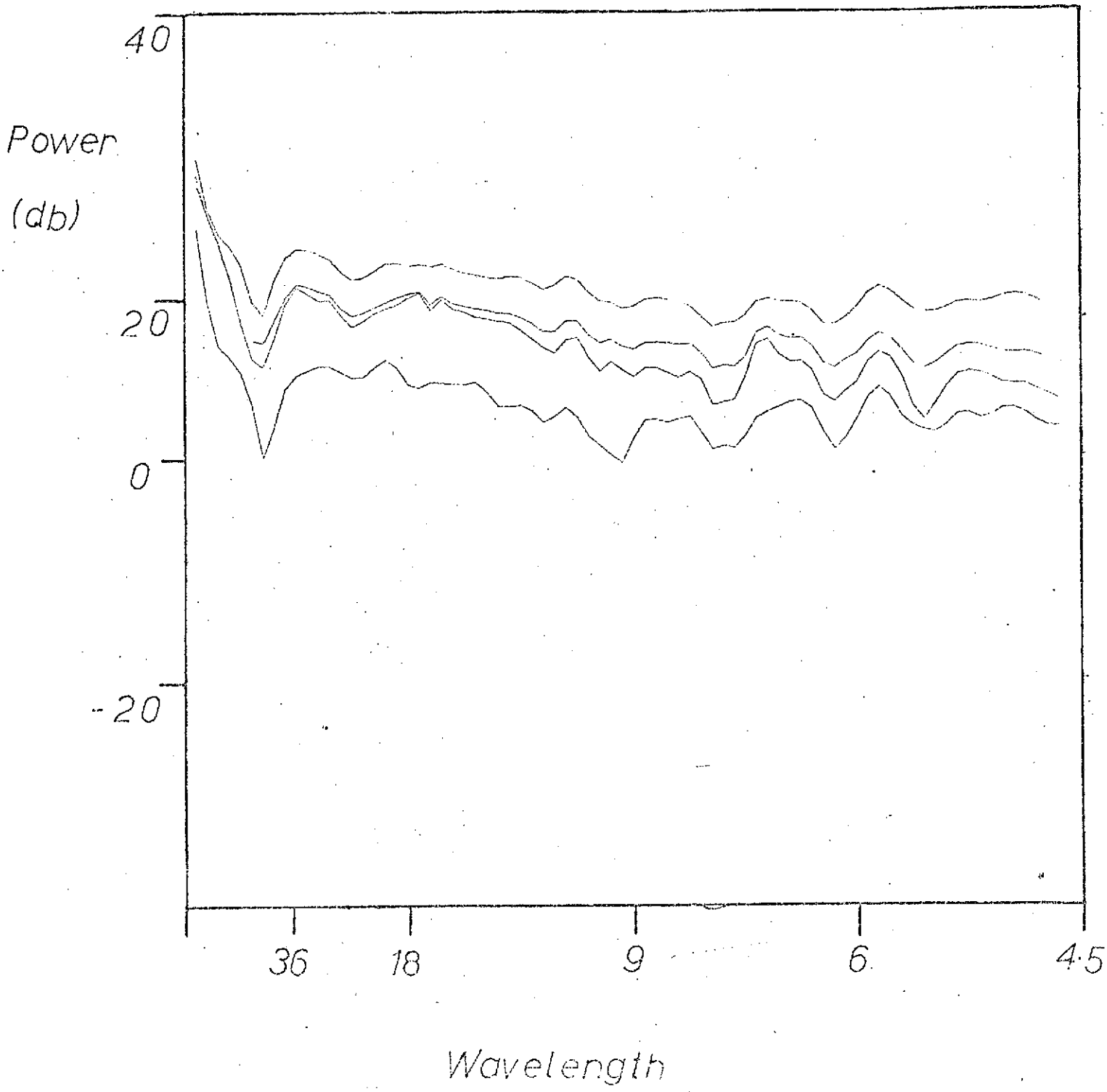

40
20
0
-20
Power
(db)
36
18
9
6
4·5
Wavelength

Figure 10. Sample output from the histogram splitting program.

DOCUMENT RESULTS , NORMLE390 , LP00 ON 20/03/74 AT 14.47

SPLITTING OF ERTS DIGITAL HISTOGRAM = WASH BAND 7

ESTIMATION OF PARAMETERS OF MIXED NORMAL DISTRIBUTIONS

40 CLASSES IN INPUT HISTOGRAM

200 DATA POINTS

MAXIMUM NUMBER OF ITERATIONS

8 STEEPEST ASCENT

10 NEWTON-RAPHSON

INITIAL ESTIMATES

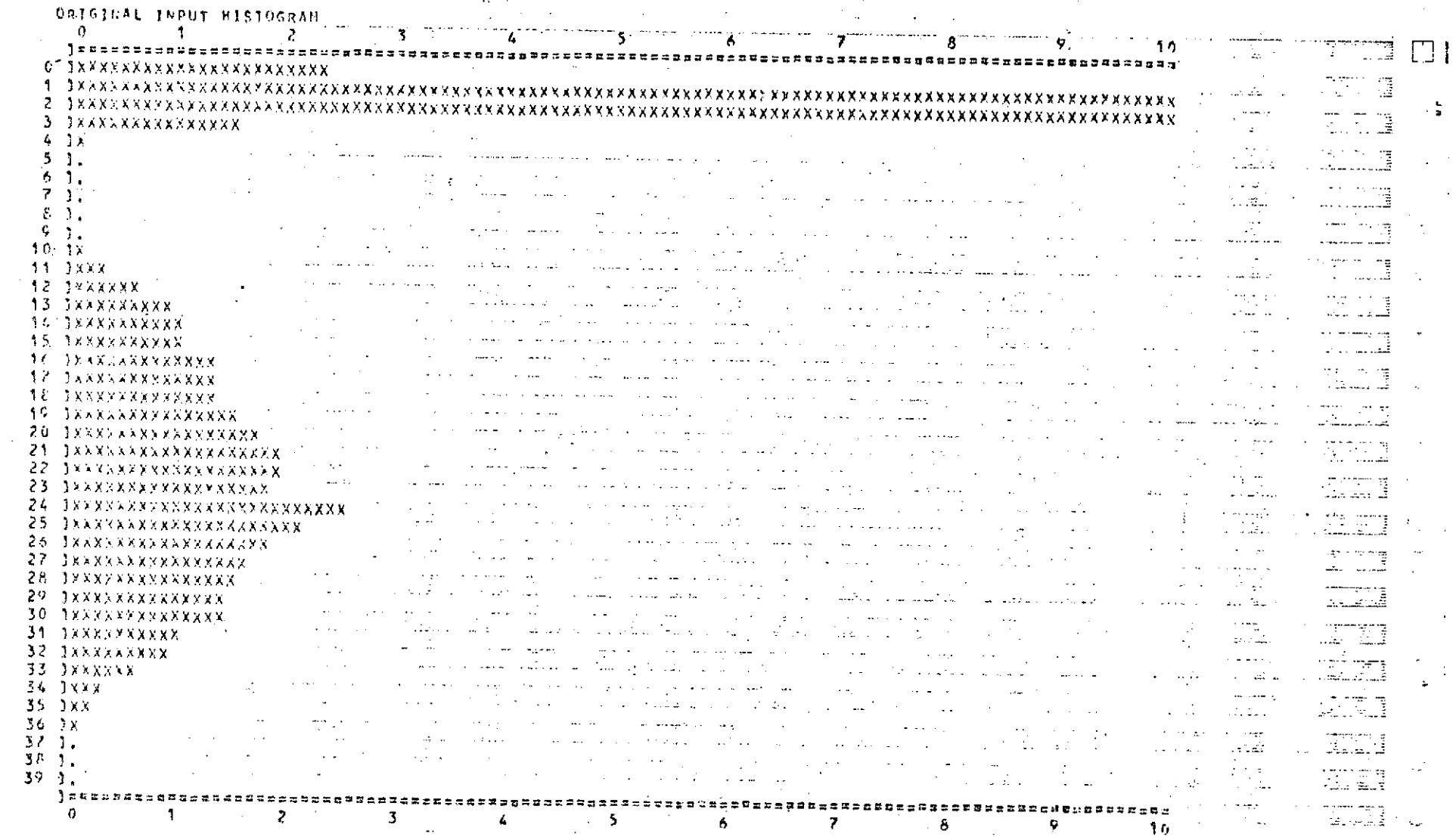
ALPHA	0.700
MU1	1.200
SIGMA1	0.500
MU2	23.000
SIGMA2	5.500

MEAN OF INPUT HISTOGRAM 8.53

STANDARD DEVIATION OF INPUT HISTOGRAM 10.6525

HISTOGRAM CONTAINS 80787 DATA POINTS

NO. OF DATA POINTS ORIGINALLY 200 ADJUSTED TO 210 BY CONVERSION OF HISTOGRAM


ORIGINAL INPUT HISTOGRAM
0 1 2 3 4 5 6 7 8 9 10
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
0 1 2 3 4 5 6 7 8 9 10

MODEL HISTOGRAM

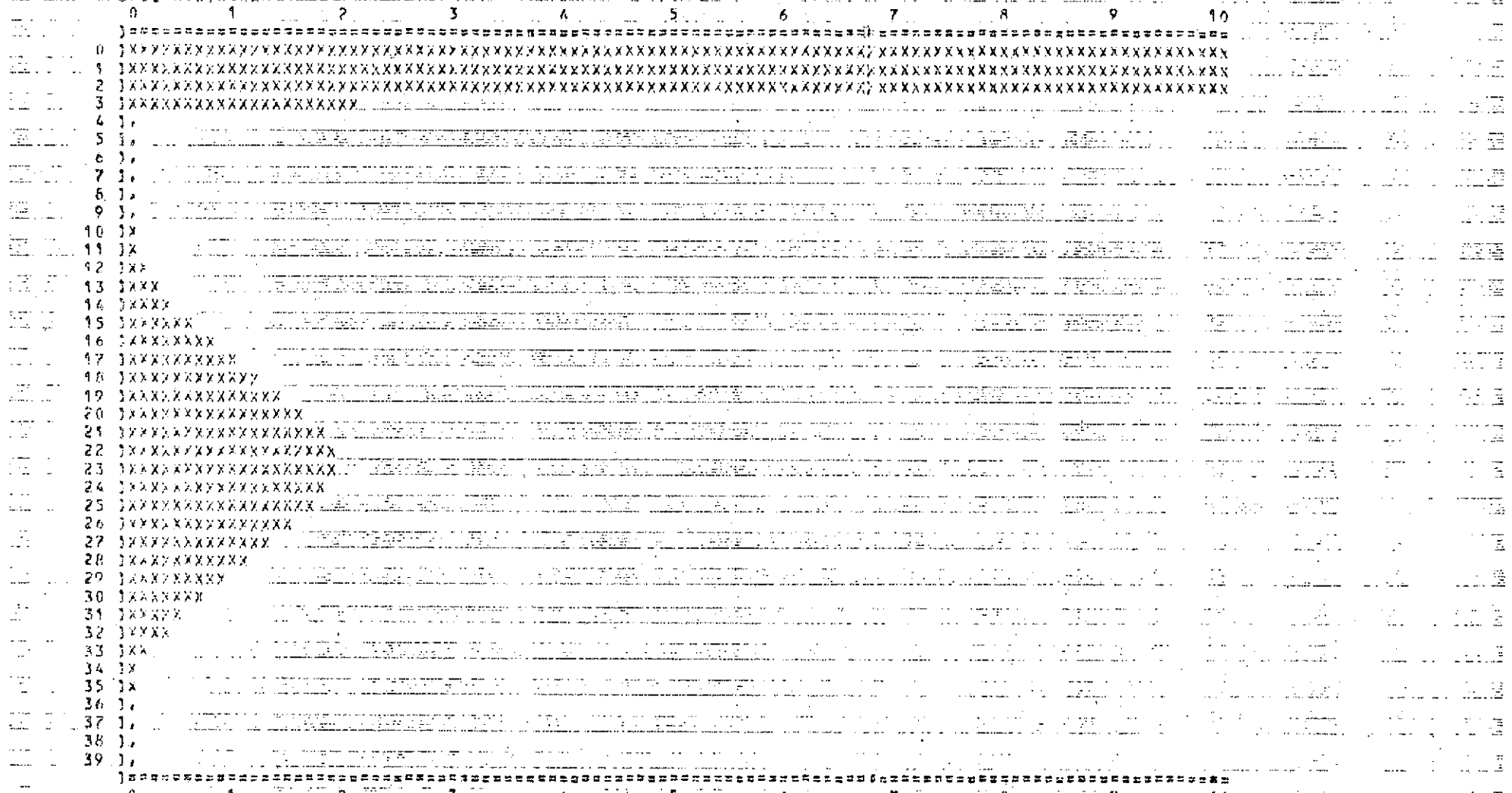


Figure 11. Dendrogram showing results of clustering the experimental set of OTU's. The error level is the square of the distance between the last two clusters amalgamated.

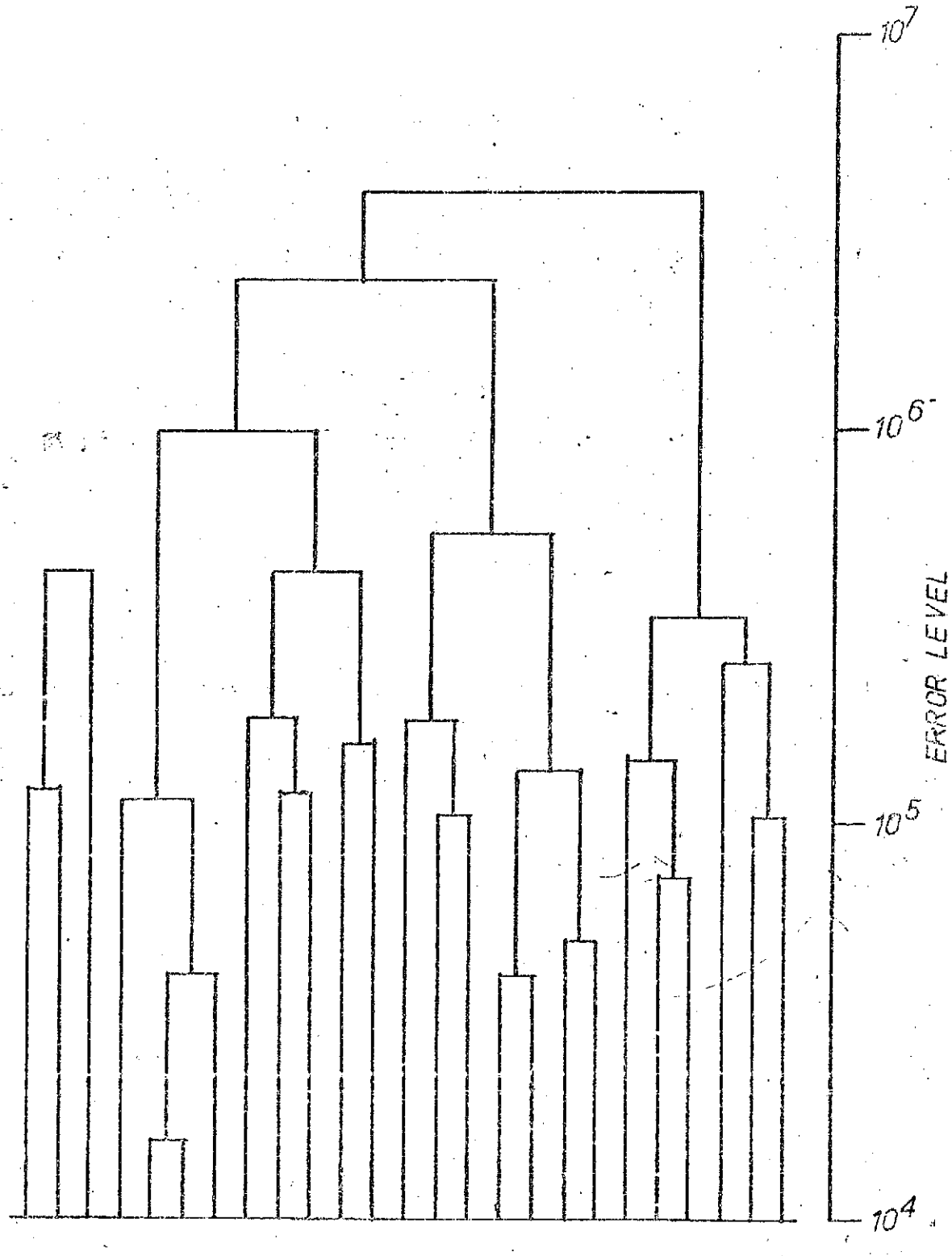
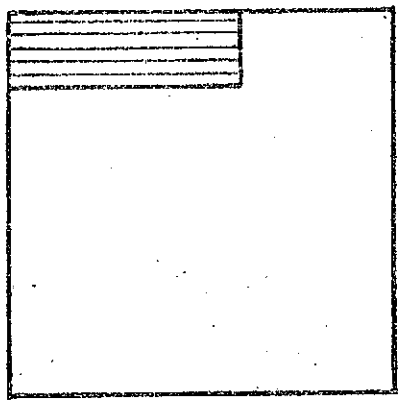
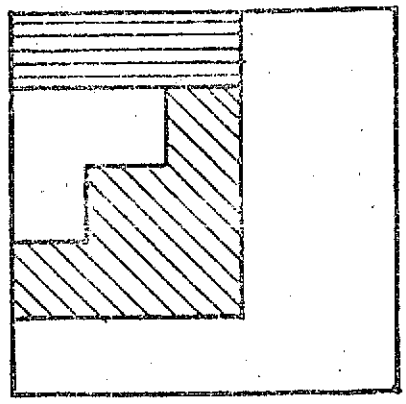

10^7
10^6
10^5
10^4
ERROR LEVEL

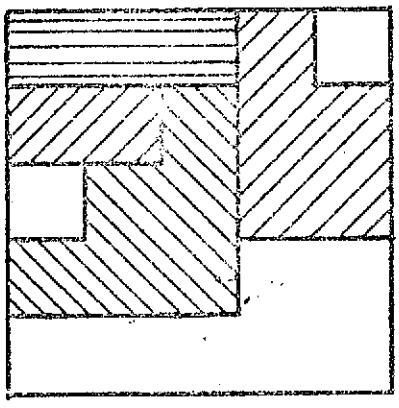
Figure 12. Maps of the groups produced in the last four steps of the clustering procedure.



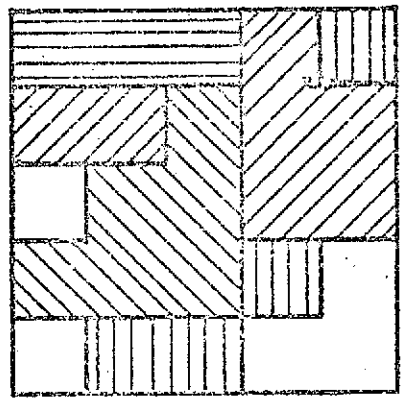
2 groups



3 groups



4 groups



5 groups

Figure 13. Graph showing the square of the distance between clusters as a function of the number of clusters.

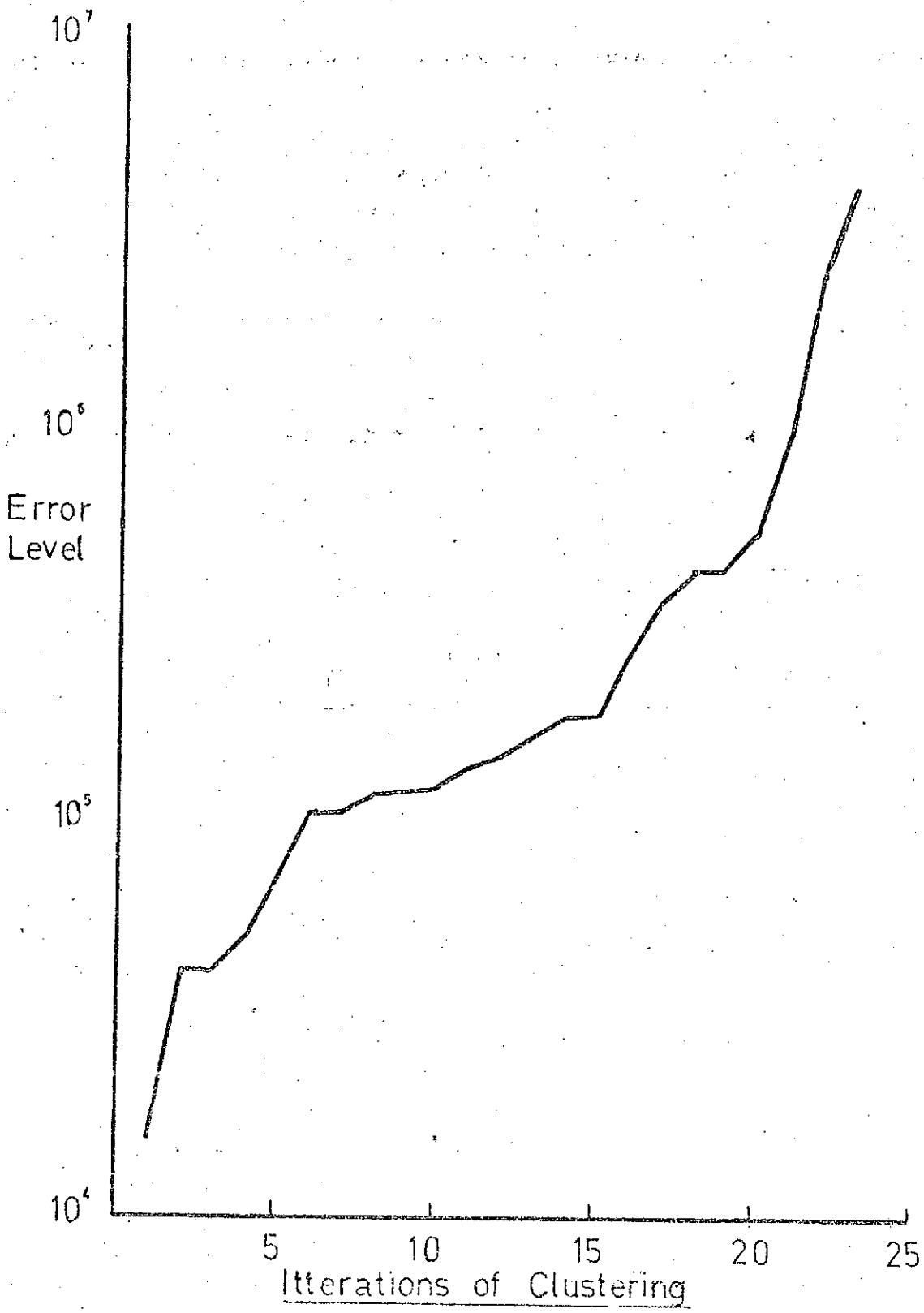
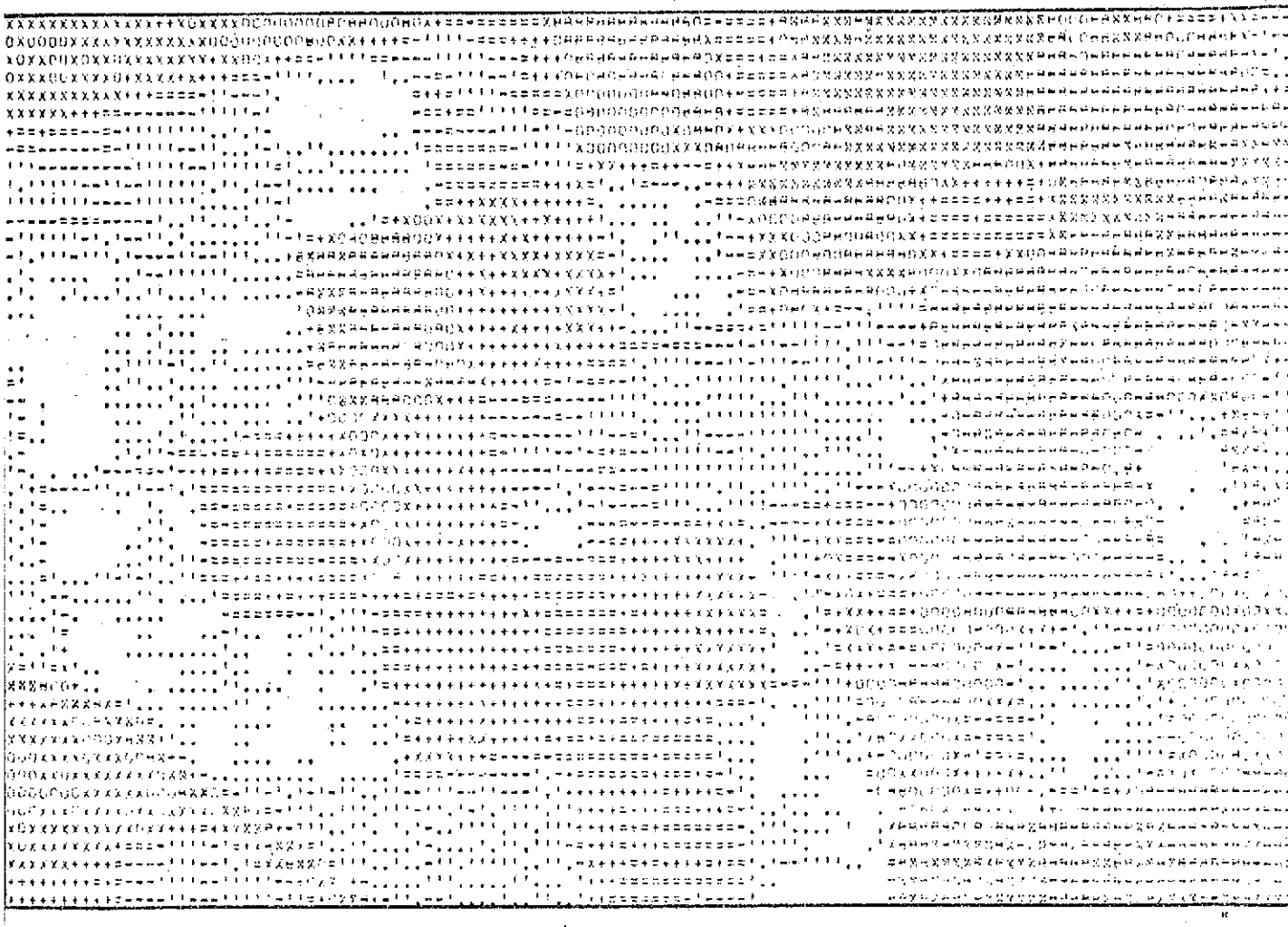
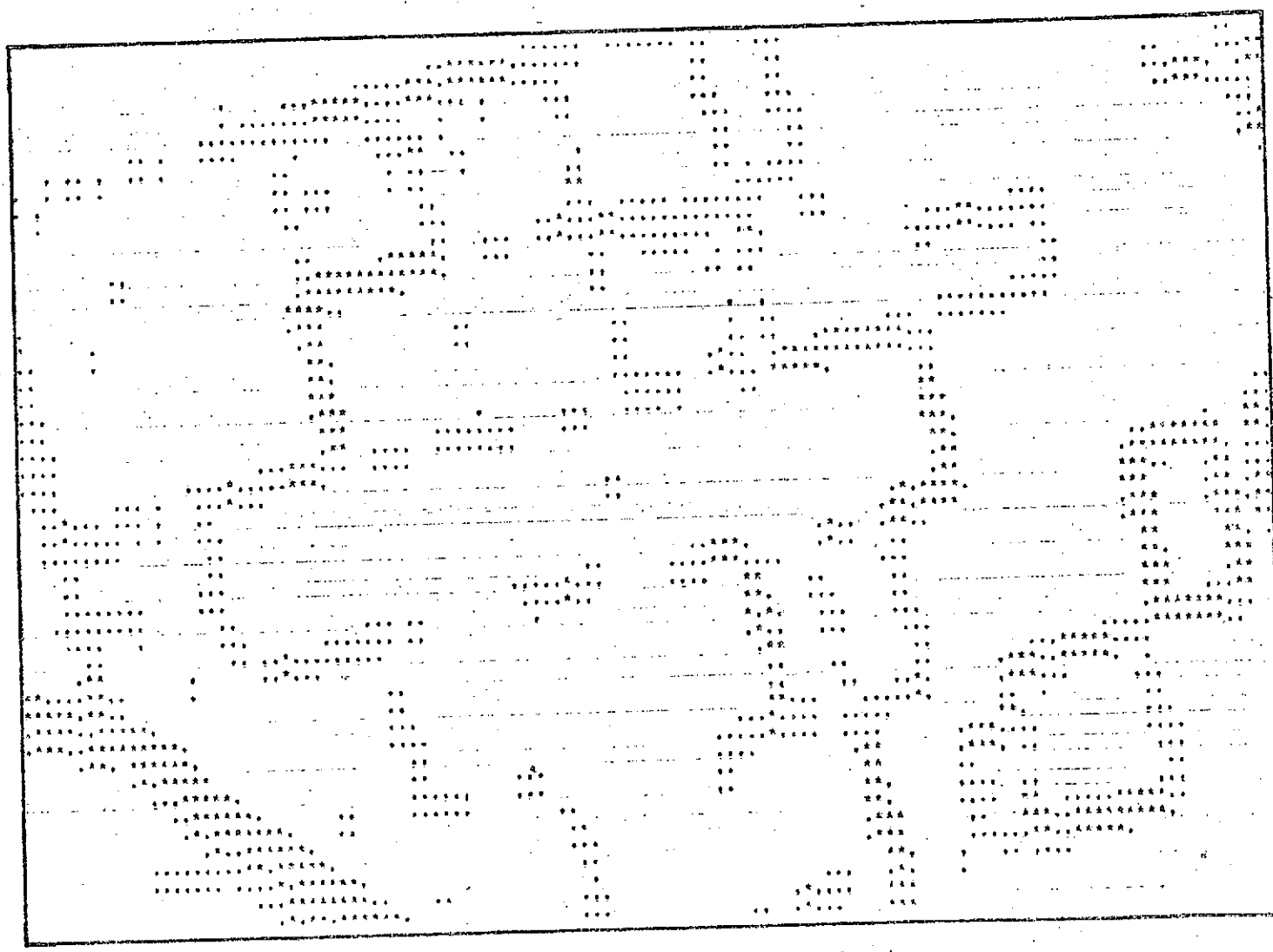

10^7
10^6
Error Level
10^5
10^4
5
10
15
20
25
Itterations of Clustering

Figure 14. A computer print out of an area in the Central Valley. The field pattern is clearly visible in this band 7 image and an irrigation canal crosses the south-west corner of the area.



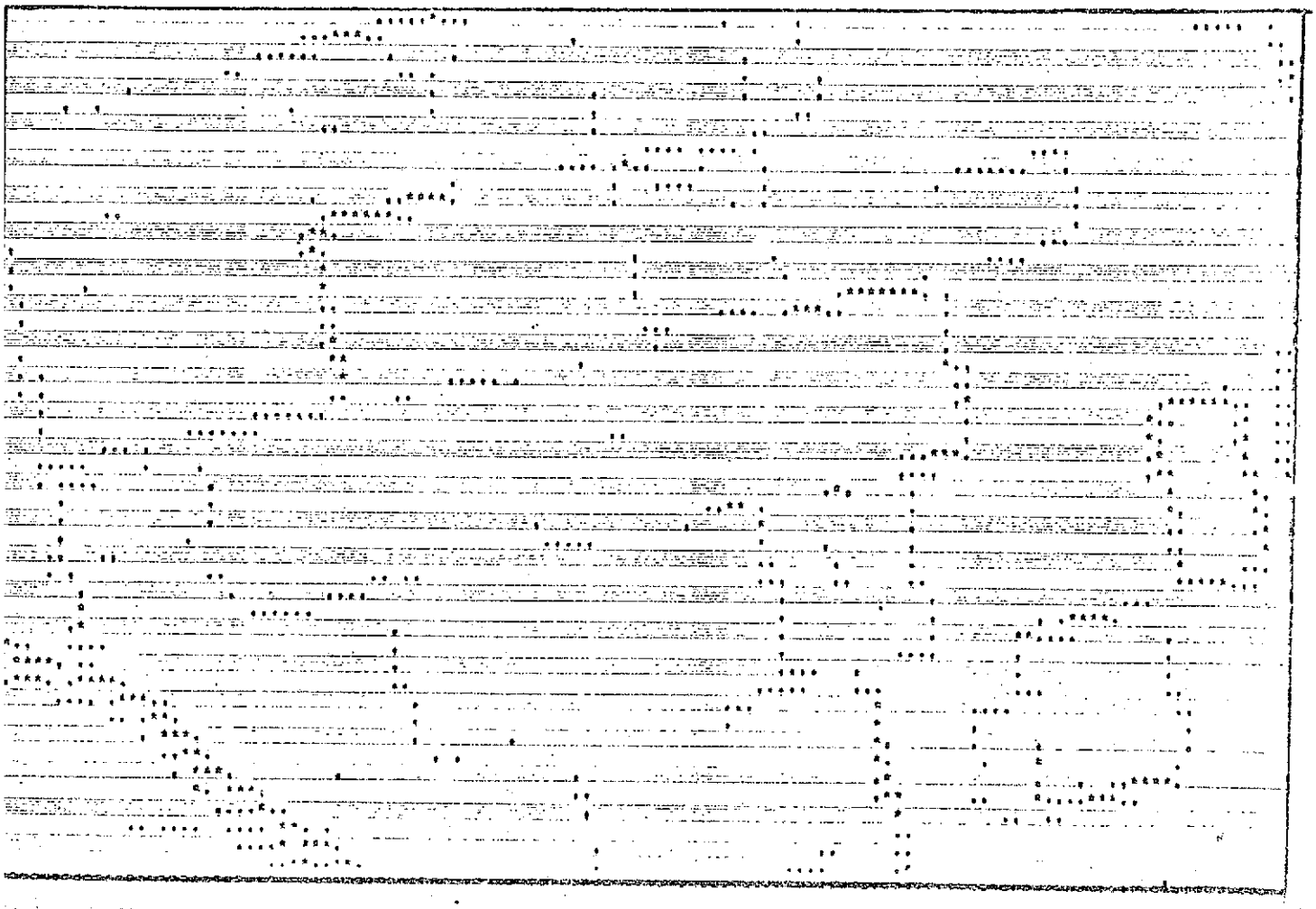
Grey scale map of area in Central Valley

Figure 15. Sharp boundaries detected in the area shown in Fig.14 using the eight nearest neighbours method. A double threshold is used so that stars represent a larger difference in gray scale value.



Field boundaries using 8 nearest pixels

Figure 16. Sharp boundaries detected using the three nearest neighbours in the north west quadrant. Note that the boundaries are more sharply defined than those in Figure 15, and that no significant features shown in Figure 15 are not also shown here.



Field boundaries using 3 nearest pixels in NW quadrant

APPENDIX I

The Generalisation of ERTS Data for Global Scale Investigations

I. E. Hill, A. C. Armstrong & K. M. Clayton

(Presented at the Symposium on European Earth Resources Satellite Experiments, February 1974)

1 INTRODUCTION

This paper describes a computer based system we, at the University of East Anglia, are developing for the analysis of imagery from the ERTS project. The starting point for our work is a belief that, although conventional air photo interpretation techniques can be, and have been, used to interpret remotely sensed data from orbital platforms, the differences between the data obtained from aircraft and from satellites warrant the development, or at least the testing, of new methods of analysis. These new methods should be designed to work in parallel with, rather than in place of, conventional analysis, with the aim of exploiting the additional information contained in the satellite imagery.

The advantages of satellite platforms over aircraft platforms are their ability to provide (i) repetitive coverage at regularly spaced points in time and (ii) global or regional coverage with a uniform data base, whereas a major disadvantage of satellite platforms is the limited spatial resolution obtainable from such altitudes. Repetitive coverage provides the opportunity for monitoring changes in surface reflectance, due to either transient phenomena such as flooding and burns, periodic effects due to seasonal changes in, for example, vegetation, or long term changes, such as urban development, forest clearance, etc. The poor resolution and the global coverage together imply that satellite imagery would be more usefully used looking for changes on a large spatial scale than changes involving areas near the limits of resolution of the sensor.

The work described here has accordingly been aimed at developing a system for monitoring satellite imagery for changes that are significant on a regional scale. Probably one of the most important uses of such a system would be as a first stage data filter for an operational earth observation satellite. There it would be used to select, for further processing, only those images which are significantly different from earlier imagery of that area.

2 TECHNIQUES

Because it is proposed that the system be used as a first stage data filter, it is important that it can make use of data that has undergone the minimum amount of pre-processing. It is probable that any satellite platform for monitoring the earth's surface on an operational and repetitive basis will return the data to earth as a digitized picture, and, for this reason, a computer based system has been chosen. Other reasons for the use of digitized data are the ability to make use both of the spectral information provided by multispectral scanners without producing colour composite photographs, a slow process when compared with the data acquisition rate, and of the full dynamic range of the sensors, without the degradation in radiometric fidelity introduced by photographic processing.

For this project large general purpose computers (an IBM 370/165 and and ICL 1903E) have been used, and the programs have been written in Fortran IV. While a large computer and a high level programming language are convenient in the experimental stages of a project such as this, it seems likely that a smaller special purpose machine, perhaps with some special handware functions, or with some analogue processing, could be more efficient for an operational system.

The first stage in the analysis of the imagery is generalisation of the data to a scale approgriate for a regional survey. Each ERTS frame contains 3×10^7 picture points for the four MSS Bands, whereas a regional map at a scale of, say, $1{:}10^6$ could show $10^2 - 10^3$ independent data points in an equivalent area. The second stage is the recognition step where the data is compared with that for previous passes over the area, and significant changes are identified. Note that these two steps in most other computer based systems for analysing remotely sensed data, are performed in the reverse order. For example, classification into surface types is usually done using the spectral signature for each picture point, and generalization to large areas (individual fields) is only made when the classification results are displayed. It is hoped that by suitable choice of the area over which this initial generalization is made, it will be possible to mask the effects of large changes taking place in a small area, and so obtain a system which is sensitive only to changes which are significant on a larger scale. For experimental purposes, Unit Test Areas (UTA's) which are square, with 50 km sides, have been chosen.

The simplest possible form of generalization is to reduce the data for each UTA to a single number for each spectral band, this number could be the mean radiance, in which case the generalization is equivalent to reducing the resolution of the sensor to match the UTA, or it could be some more complex parameter. For most purposes, this would be to drastic a reduction in the information content of the data, although this level of generalization could possibly be used to recognise areas with a high degree of cloud cover.

Comparison of the histograms can be done visually or by means of statistical tests. In the latter case the Chi-squared and Kolmogorov Smirnov tests have been used.

3. PRELIMINARY RESULTS

In the previous sections the aims and lines of investigation of the project have been outlined. Regrettably it is still too early in the experiment to say how effective this technique will be, and in the remainder of this paper some early results and problems will be discussed.

The area which has been chosen as a test site is covered by one ERTS frame, and is shown in Figure 1. The area covers part of the Central Valley in California and extends as far as the coastal ranges in the south-west. A large area of the Sierra Nevada range is included on the eastern side of the frame. Digital data on computer compatible tapes has been obtained for two successive passes over the area in later summer 1972 and a third set of data has been obtained for spring 1973. In addition a second test area in central England was selected, but repetitive coverage is not yet available for this area.

One essential requirement of the technique is that the gray scale distribution in unchanged areas is sufficiently repeatable for changes in other areas to be recognised. Figure 2 shows the histograms for all four bands for a UTA containing predominantly agricultural land in the Central Valley. The only changes visible on photographic prints of the imagery relate to a few fields which together make up 1 per cent or less of the UTA. The histograms have a similar shape for both scenes 1038-18114 and 1056-18114, but there is a small shift to lower grey scale values for all bands for the later pass. This shift can be accounted

for by the decrease in solar irradiance associated with the decrease in solar elevation. Table 1 shows that the ratios of mean radiances is comparable with the ratio of the sine of the elevation angle.

Although there is very good agreement between the large scale structure of the histograms, particular for band 7, there are many smaller features which are not duplicated. The histograms for bands 4, 5 and 6 tend to be "spikey", and even in the band 7 histogram there is a small spike near the peak of the distribution. These spikes are not real structure in the gray scale frequency distribution, but originate from the mapping of the sensor gray scale with its 64 values to the CCT gray scale which has 128 levels for bands 4, 5 and 6, and 64 levels for band 7. Because of the non-linear transfer function and expanded CCT gray scale for bands 4, 5 and 6, and because of the radiometric corrections applied to all bands during the transfer, there is not a one-to-one mapping of sensor gray levels to CCT gray levels. This is illustrated by Figure 3 which gives the gray scale frequency distributions for successive groups of six scan lines down a strip one quarter of the image in width. The features to be noted are the blank columns corresponding to 2 gray scale levels for this band 5 data which continue through the distribution for several successive lines. The calibration process changes the mapping from sensor levels to tape levels as the sensitivity of the sensor changes, and over a sufficiently large number of lines this tends to smooth the frequency distribution . Alternately, the histograms may be smoothed, by use of a suitable weighting function, to give about 32 independent gray levels. However the smoothing function for bands 4, 5 and 6 is dependent on the actual gray level and results in a non-linear relationship between tape gray scale and radiance. The effect of the non-linearity is to weight changes in low reflectance surface cover more heavily than changes in

the upper half of the radiance scale.

One part of the computer system which has been extensively tested is the segment which aligns the digital data relative to the UTA. The confidence that can be placed in changes in surface reflectance detected by this system depends on the accuracy of this alignment. For example with 50 km square UTA's a r.m.s. positional error of 1 km places a lower limit on the fraction of the area for which changes can be reliably detected, of about 5% of the total area of the UTA. The proposed computer system uses the latitude and longitude grid information in the annotation record of the CCT to convert the coordinates of the corners of the UTA to tape coordinates (scan line, picture element). The pre-launch estimate of the accuracy of this conversion was about 500 metres r.m.s. for the three scenes for which tests have been made, the errors are in the range 5-10 km. Fig.4 shows the positions found for the dam of Honey Lake, California, from the annotation data for scenes 1038-18114 and 1056-18114. This error is sufficiently large to obscure any changes involving less than about one quarter of the area of the UTA, and so, for most purposes, is large enough to prevent this monitoring technique working.

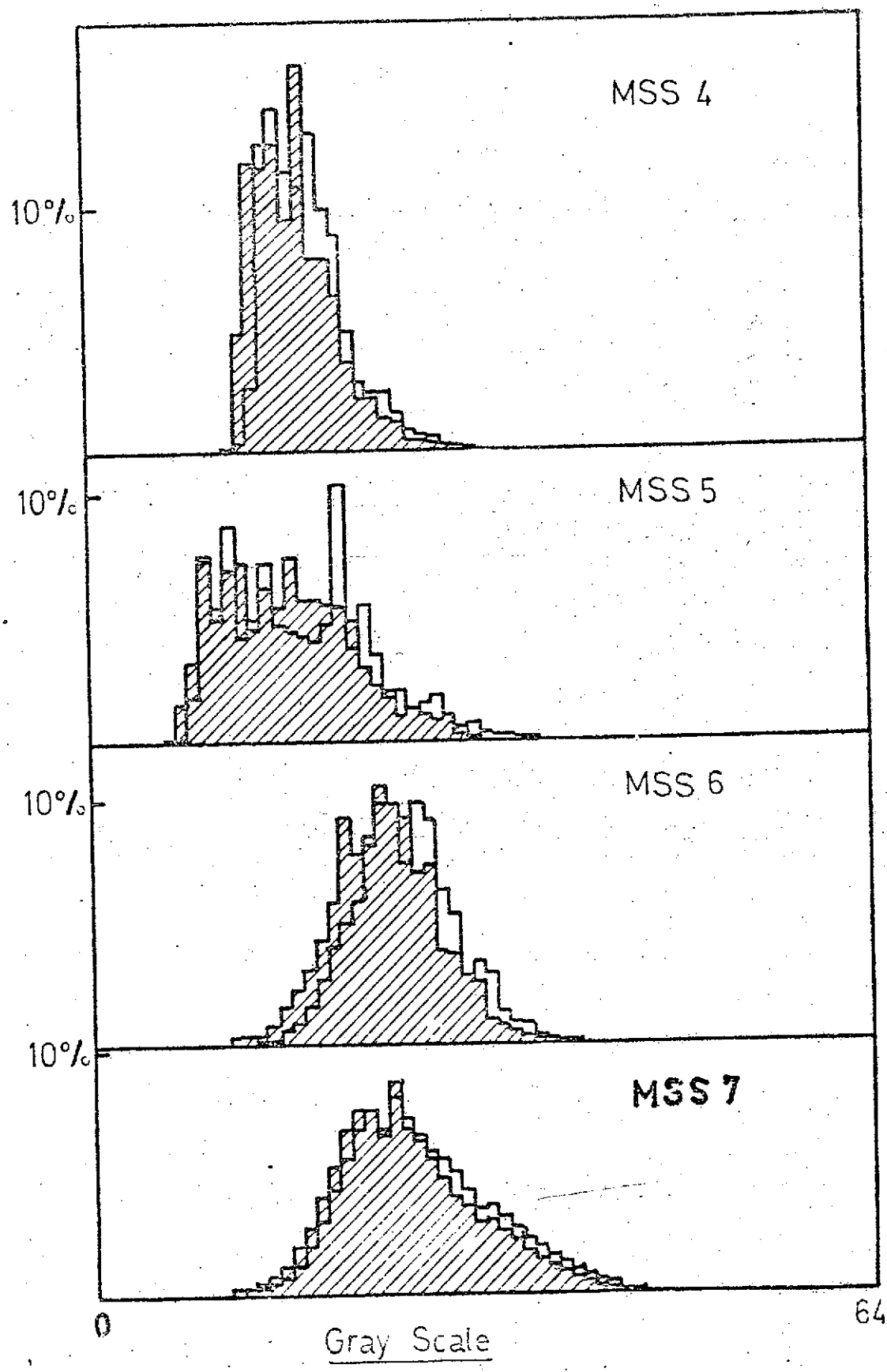
The UTA could of course be more accurately positioned relative to the digital data by use of several control points on the ground, but this has the disadvantage that, for an operational system working on a world wide scale, the number of control points would soon become unmanageable. The ideal solution for an operational satellite platform would be improved attitude and position information and control, but this might be technically not feasible. One alternative method, which is being tested,

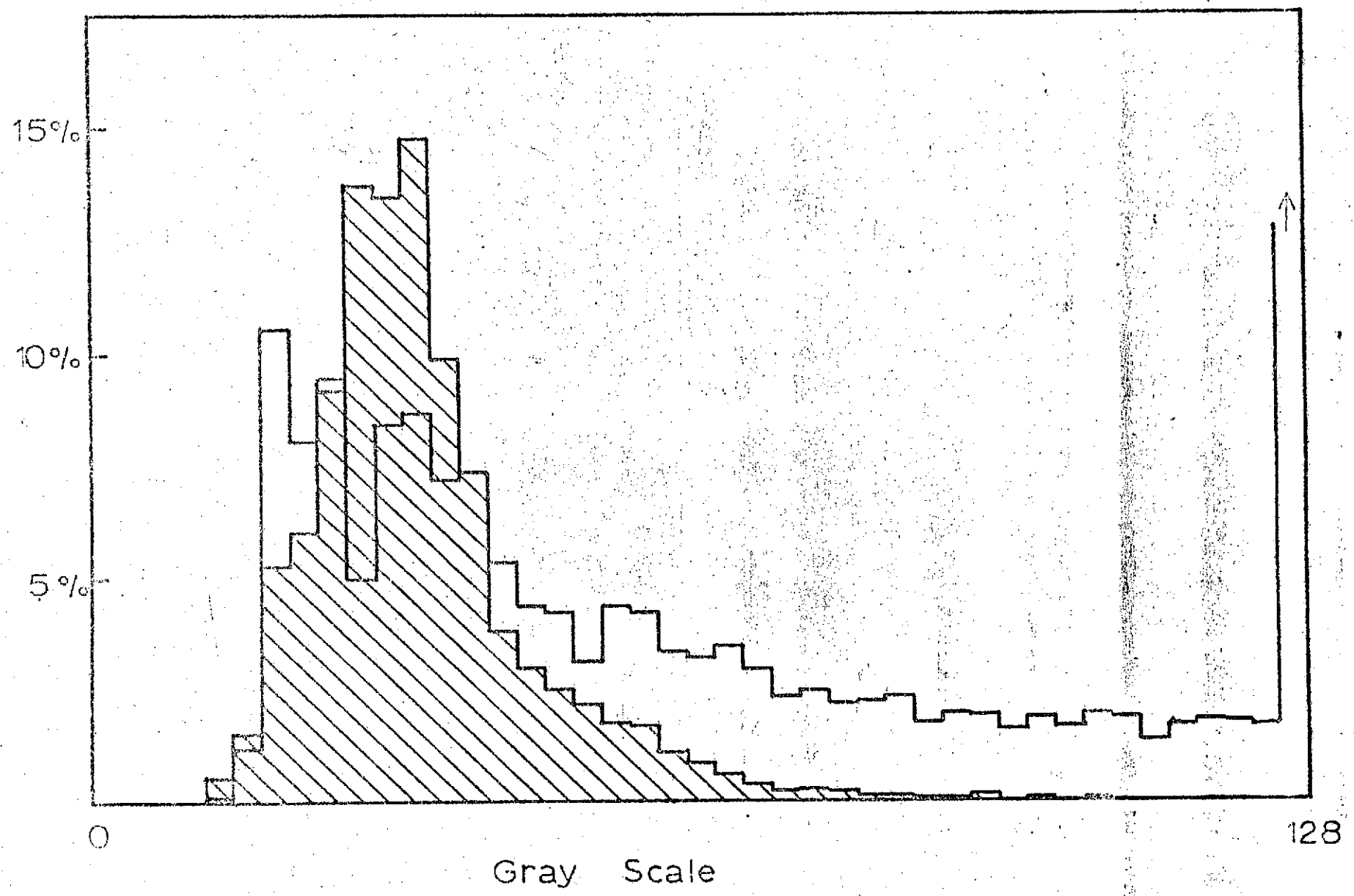
is to align one image with the ground by manual methods, and then to use this primary image as a correlation mask for successive passes.

As a final example of the sort of difficulties that need to be overcome before the computer system can be used operationally, Figure 5 shows the effects of clouds and cloud shadow on the gray scale distribution. The histograms are for the same part of the Sierra Nevada range, but in one case the area is partially cloud covered. This produces a sharp peak at high radiance, as expected, but in addition the cloud shadow, and structure within the cloud, are sufficient to make the remainder of the histograms totally different. The implication of this is that changes in the cloud free area could only be detected if some spatial information is used to remove the whole cloud affected area from both images before preparing the histograms.

4 CONCLUSIONS

The results presented above have tended to emphasize the difficulties that have been met in developing the computerized system, and have given a rather negative outlook on the usefulness of the technique. At this stage however, there is still much development required, and no tests of the effectiveness of the system will be possible before difficulties such as those described above have been solved. Other work proceeding in parallel with that described here, does however, give some indication that using the gray scale histograms as a method of generalizing the data is a satisfactory solution. In particular, it seems possible to use the histograms alone as a means of dividing the image into surface types using standard procedures.


MSS 4
10%
MSS 5
10%
MSS 6
10%
10%
MSS 7
0
64
Gray Scale


15%
10%
5%
0
128
Gray Scale